Dicembre 1990
Numero 268
Anno XXIII
Volume XLV

LE SCIENZE
edizione italiana di SCIENTIFIC AMERICAN

Il calcolatore Grande Maestro di scacchi

Nei 40 anni trascorsi dai primi tentativi di programmare calcolatori capaci di giocare a scacchi, le macchine hanno battuto in torneo giocatori di tutti i livelli di abilità: il campione del mondo sarà il prossimo a cadere?

di Feng-hsiung Hsu, Thomas Anantharaman, Murray Campbell e Andreas Nowatzyk

Nel gennaio 1988, durante una conferenza stampa tenuta a Parigi dal campione del mondo di scacchi Gary K. Kasparov, un giornalista chiese se un calcolatore sarebbe stato in grado di sconfiggere un grande maestro prima della fine del secolo. Senza indugio Kasparov rispose: «Certamente no, e se qualche Grande Maestro avesse davvero difficoltà a giocare contro un calcolatore sarei felice di dargli qualche consiglio».

Dieci mesi dopo questa affermazione di Kasparov, in un importante torneo tenutosi a Long Beach, in California, il Grande Maestro Bent Larsen, che era stato in lizza per il titolo mondiale, fu sconfitto da un calcolatore da noi ideato alla Carnegie-Mellon University. Oltre a questo incontro, *Deep Thought*, una combinazione di software e di hardware dedicato, vinse altre cinque partite, ne pareggiò una e una ne perse, conquistando in tal modo il primo posto *ex aequo* con il Grande Maestro Anthony Miles. Dato che per regolamento un calcolatore non può vincere denaro in un torneo, Miles intascò tutti i 10 000 dollari del primo premio. (Tuttavia Deep Thought lo sconfisse un anno dopo in una partita di esibizione.)

Entro l'estate 1990 - durante la quale tre fondatori del gruppo Deep Thought sono entrati a far parte della IBM - il calcolatore aveva vinto cinque delle 10 partite disputate in torneo contro Grandi Maestri e 12 delle 14 partite giocate contro Maestri Internazionali. Alcuni di questi incontri e decine di altri che hanno opposto il calcolatore ad avversari di livello inferiore sono stati giocati sotto gli auspici della US Chess Federation, che ha attribuito a Deep Thought una valutazione Elo di 2552. [È da notare tuttavia che la scala statunitense sopravvaluta i punteggi di circa 100 punti rispetto a quella della FIDE (Fédération Internationale des Èchecs).] Questa valutazione pone Deep Thought fra i Grandi Maestri di livello più basso, ma è comunque molto superiore a quella di un giocatore medio di torneo, che in genere ha un punteggio intorno a 1500 (*si veda l'illustrazione alle pagine 14 e 15*). Nelle partite disputate dopo l'agosto 1988, quando il calcolatore è stato portato alla sua attuale velocità di analisi di 750 000 posizioni al secondo, la sua valutazione Elo è stata superiore a 2600.

Il modello perfezionato del nostro calcolatore, che si prevede disputerà le sue prime partite nel 1992, sarà dotato di un hardware molto più potente che aumenterà di oltre 1000 volte la velocità di analisi, fino a circa un miliardo di posizioni al secondo. Questa sola modifica potrebbe rendere il discendente di Deep Thought addirittura più forte di Kasparov o di qualunque altro giocatore della storia.

Ma perché si insegna a una macchina come minacciare un re di legno su una scacchiera? In primo luogo gli scacchi vengono tradizionalmente considerati in Occidente come il principale gioco di intelligenza; per dirla con Goethe, sono «la pietra di paragone dell'intelletto». Secondo alcuni, la costruzione di un calcolatore in grado di giocare bene a scacchi dimostrerebbe che è possibile mettere a punto modelli dei processi di pensiero o, viceversa, che in questo gioco non è necessario pensare. Sia l'una sia l'altra conclusione imporrebbero di rivedere ciò che si intende comunemente per intelligenza.

Oltre a ciò, un calcolatore che giochi a scacchi costituisce un problema tecnico appassionante, come scrisse 40 anni fa su «Scientific American» Claude E. Shannon, il padre della teoria dell'informazione: «Si studia il problema del gioco degli scacchi al fine di sviluppare tecniche che possano essere sfruttate per applicazioni più pratiche. Il calcolatore che gioca a scacchi rappresenta un punto di partenza ideale per diverse ragioni. Il problema è definito nettamente, sia nelle operazioni permesse (le mosse degli scacchi) sia nell'obiettivo finale (lo scacco matto); inoltre, non è tanto semplice da essere banale né tanto complesso da rendere impossibile una soluzione soddisfacente. E organizzando una partita fra una macchina del genere e un giocatore si può avere una chiara indicazione della capacità del calcolatore in questo tipo di ragionamento.»

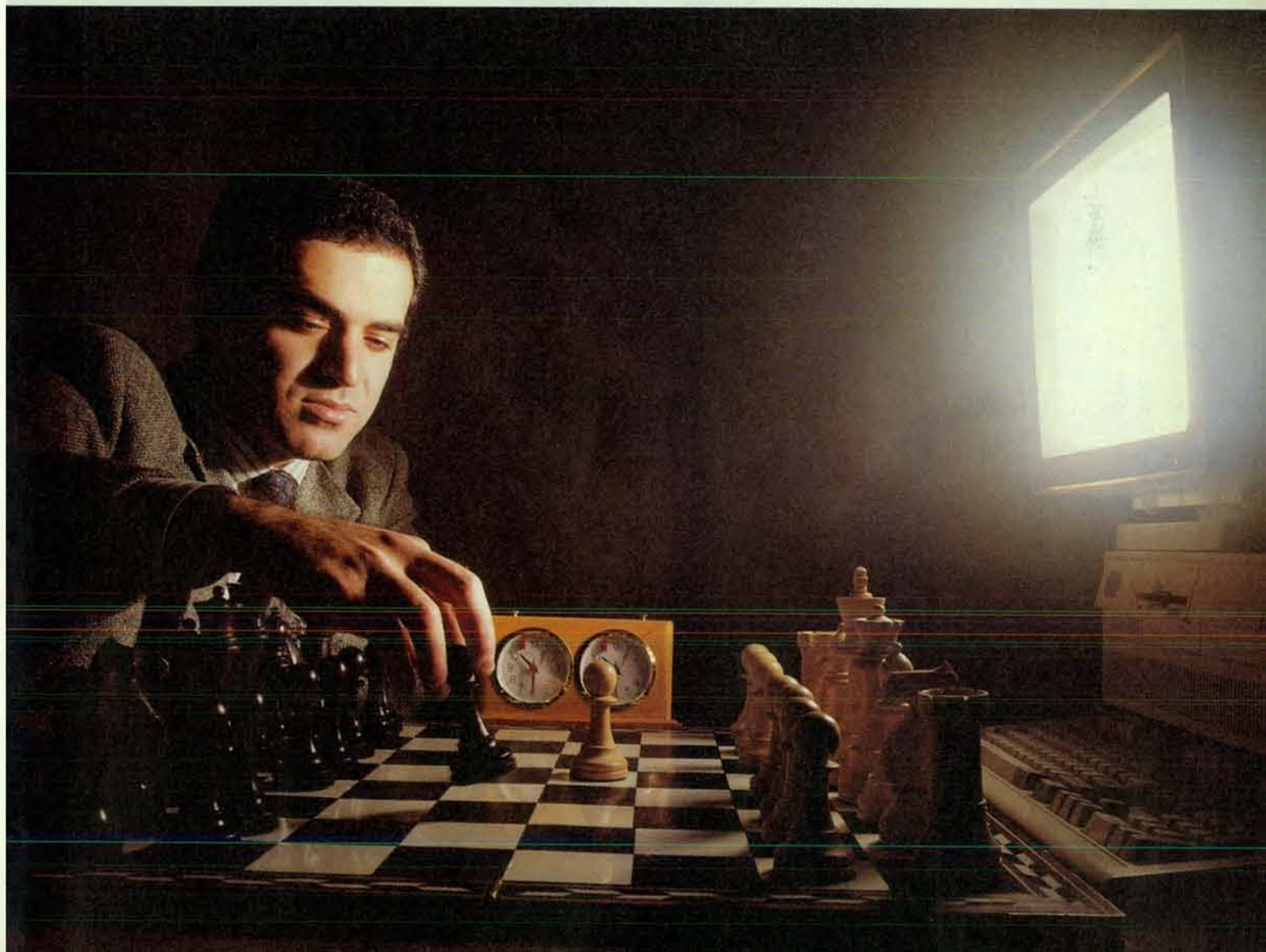
Forse la conseguenza pratica più importante di questi studi è stata la dimostrazione dell'efficacia dell'analisi al calcolatore; il perfezionamento delle relative tecniche porterà senza dubbio a progressi nei settori della progettazione di reti, della messa a punto di modelli di reazioni chimiche e anche dell'analisi linguistica.

La storia delle macchine che giocano a scacchi iniziò negli anni successivi al 1760, allorquando il barone Wolfgang von Kempelen presentò in Europa il suo automa, che venne soprannominato «il turco» perché si presentava come una marionetta con tanto di mustacchi e turbante, apparentemente azionata da un complicato meccanismo che si trovava in un compartimento sottostante. In genere il congegno giocava bene e una volta mandò su tutte le furie Napoleone Bonaparte sconfiggendolo in 19 mosse. Edgar Allan Poe, fra gli altri, scoprì in seguito il segreto dell'automa: nel compartimento era nascosto un abile scacchista di piccolissima statura. Tuttavia la ragione che fece intuire a Poe l'imbroglio era erronea: lo scrittore affermò infatti che se si fosse trattato di una macchina autentica, non avrebbe mai dovuto perdere!

Alan M. Turing, matematico, studioso di calcolatori e crittografo, fu tra i primi a confrontarsi con il problema di un calcolatore che fosse in grado di giocare a scacchi, ma si accorse che era più semplice attuare il suo programma di generazione delle mosse e di valutazione della posizione a mano anziché con la macchina. Il tedesco Konrad Zuse e altri studiosi fecero tentativi analoghi, ma l'opera fondamentale fu quella di Shannon. Egli prese le mosse dalle scoperte di John von Neumann e Oskar Morgenstern che, nella loro teoria generale dei giochi, avevano escogitato un algoritmo «minimax» mediante il quale era possibile determinare la mossa migliore.

Essenzialmente, l'algoritmo determina tutte le possibili posizioni risultanti dopo un certo numero di mosse dei due giocatori, assegna loro un valore numerico e, in funzione di questi valori, stabilisce quale sia la mossa migliore. All'inizio il generatore di mosse calcola tutte le mosse possibili per il calcolatore a partire dalla posizione in corso, poi tutte le possibili risposte dell'avversario e così di seguito. L'algoritmo esamina pertanto un albero di gioco, nel quale ciascuna posizione è connessa alle posizioni successive consentite dalle regole del gioco.

A ciascun livello dell'albero vi è un numero di posizioni circa 38 volte superiore a quello del livello precedente (38 è il numero medio di mosse possibili in una tipica posizione di gioco), oppure circa sei volte superiore se si utilizza la «potatura alfa-beta» (*si veda la «finestra» a pagina 16*). La maggior parte delle posizioni si trova dunque alle estremità dei rami dell'albero, dove esso continua a crescere fino a quando l'analisi non sia completata o il tempo a disposi-



Il campione del mondo Gary Kasparov posa con un IBM PS/2, usato per comunicare con Deep Thought, prima di una partita disputata alla fine del 1989. Kasparov vinse facilmente sebbene la macchina venga considerata al livello di un Grande Maestro.

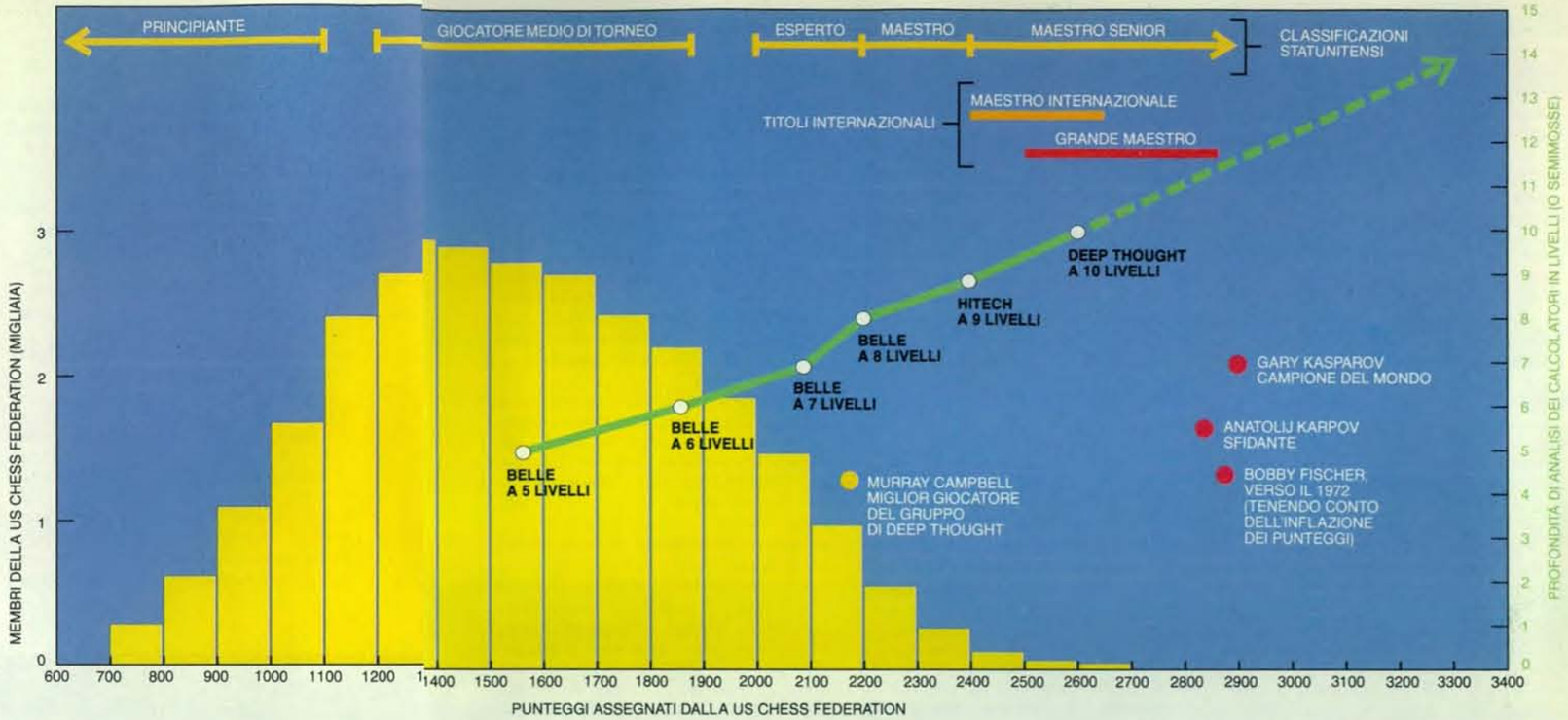
zione del calcolatore non sia scaduto. Una funzione di valutazione assegna quindi un punteggio a ciascuna posizione finale, attribuendo per esempio un 1 a uno scacco matto ai danni dell'avversario, un − 1 a uno scacco matto ai danni della macchina e uno 0 a una patta. È anche possibile considerare e confrontare situazioni meno nette: per esempio il calcolatore può tener conto della qualità (ossia il valore dei pedoni e delle figure) e calcolare il valore delle posizioni in base a parametri che definiscono la disposizione delle figure, lo schieramento dei pedoni, l'occupazione di una colonna aperta, il controllo del centro e così via.

Si può migliorare il gioco di un calcolatore aumentandone la profondità d'analisi (ossia il numero di semimosse che esso considera prima di valutare le posizioni) o perfezionandone la funzione di valutazione. La macchina giocherebbe in maniera perfetta se fosse in grado di esaminare in modo esaustivo tutte le possibili successioni di mosse fino al matto di una delle due parti o alla patta. Una macchina siffatta potrebbe allora sorprendere l'avversario annunciando già alla prima mossa: «Il bianco muove e vince in 137 mosse» oppure giudicando la posizione insostenibile e abbandonando. Ma, se l'analisi esaustiva è facile in giochi come il filetto, è impossibile negli scacchi, dove si possono avere 10^{120} partite diverse. Un gioco egualmente perfetto potrebbe nascere dall'esame di un solo livello dell'albero, ammesso che la valutazione delle posizioni fosse di livello pari a quella rivendicata da Capablanca, il quale affermava di prevedere una sola mossa: la migliore.

I primi studiosi a occuparsi di questo settore erano ben lungi dall'avanzare pretese del genere, dato che fino al 1958 non si era neppure in grado di programmare macchine che rispettassero le regole degli scacchi. Dovevano infatti passare altri otto anni prima che il programma MacHack-6, ideato da Richard D. Greenblatt del Massachusetts Institute of Technology, riuscisse a raggiungere il livello del giocatore medio di torneo.

Gli autori di programmi per gli scacchi, via via più numerosi, cominciarono a dividersi in due fazioni dalle filosofie distinte, che potremmo chiamare degli «strateghi» e dei «tecnici». I primi affermavano che un calcolatore dovrebbe giocare a scacchi allo stesso modo di una persona, decidendo le mosse per mezzo di ragionamenti espliciti, mentre i tecnici adottavano un atteggiamento meno rigido, sostenendo che un tipo di ragionamento valido per gli esseri umani potrebbe non essere applicabile a una macchina. Agli inizi, quando la progettazione di calcolatori capaci di giocare a scacchi era più una questione teorica che pratica, gli strateghi sembrarono avere il sopravvento.

La competitività sempre crescente dei calcolatori è chiara in questo confronto fra i punteggi dei 35 000 membri della US Chess Federation e quelli di macchine con diverse profondità di analisi (*in verde*). I calcolatori guadagnano circa 200 punti per ciascuna semimossa, o livello dell'albero di gioco, in più che esaminano. Attualmente Deep Thought analizza 10 livelli e ha una valutazione di circa 2600; il suo successore potrà esaminare 14 o 15 livelli e ottenere così una valutazione di gran lunga superiore.



La fazione dei tecnici divenne protagonista negli anni settanta, quando si scoprì che le prestazioni dei programmi erano quasi proporzionali alla profondità di analisi: a ogni livello aggiuntivo dell'albero di gioco che viene esaminato l'Elo aumenta di 200 punti (*si veda l'illustrazione in queste due pagine*). I programmatori si affannarono allora per avere accesso a calcolatori sempre più veloci e applicarono tutti i trucchi del mestiere per ottenere una maggiore profondità d'analisi dalla potenza di calcolo disponibile.

Tuttavia, mettere a punto la ricerca delle posizioni non basta. I primi programmi per gli scacchi generavano le posizioni senza alcun discernimento: consideravano, per esempio, distinte successioni che, per inversione di due mosse, portavano alla medesima posizione. Queste operazioni ridondanti vengono attualmente evitate conservando le posizioni in matrici di memoria. Questo stratagemma facilita anche il processo di potatura dell'albero di gioco, in quanto consente di scartare immediatamente numerose successioni di mosse prive di interesse.

Il problema più arduo è quello di stabilire fino a quale profondità occorra esaminare una diramazione dell'albero di gioco. Non è certo possibile continuare l'analisi indefinitamente, ma sarebbe comunque auspicabile non interromperla in una posizione «instabile», come quella che si ha nel bel mezzo di uno scambio di pezzi. Si supponga, per esempio, che il calcolatore esamini otto semimosse per ciascuna diramazione e arrivi a una posizione, dopo l'ottava semimossa, in cui esso apparentemente guadagna un cavallo in cambio di un pedone. Anche se con la mossa immediatamente successiva l'avversario potesse recuperare il cavallo e ottenere il guadagno netto di un pedone il calcolatore perseguirebbe ostinatamente il vantaggio illusorio.

Questo «effetto orizzonte», come viene chiamato, spinge spesso i calcolatori a commettere una sorta di suicidio scacchistico che non si osserva neppure in partite tra giocatori alle prime armi: all'improvviso e senza alcuna ragione apparente, la macchina comincia a sacrificare pedoni e figure, riducendosi in breve in una posizione insostenibile. Per diminuire la probabilità di errori di questo tipo, virtualmente tutti i programmi attuali comprendono una fase aggiuntiva di ricerca delle posizioni stabili, il cui scopo è esaminare una sequenza di cattura di pezzi fino al raggiungimento di una posizione stabile che consenta di applicare in un contesto non fuorviante la funzione di valutazione.

Negli anni settanta e agli inizi degli anni ottanta si assistette al predominio delle macchine che si basavano sulla «forza bruta», nel senso che il loro successo era dovuto a un'abile attuazione delle strategie di ricerca e analisi delle posizioni descritte in precedenza. Questo periodo fu dominato quasi ininterrottamente dal programma Chess 4.0 della Northwestern University e dai programmi 4.X da esso derivati. Via via adattati alle nuove generazioni di calcolatori, questi programmi sono andati migliorando costantemente il loro punteggio nella scala di valutazione fino a superare nel 1979 il livello di Esperto (2000 punti).

Sempre negli anni settanta si sono avuti diversi tentativi di costruire calcolatori dedicati per il gioco degli scacchi. Il più famoso, Belle, realizzato presso gli AT&T Bell Laboratories, raggiunse il livello di Maestro (2200 punti) nel 1983. L'apogeo dell'era delle macchine che si basavano sulla forza bruta si ebbe nel 1986 con l'emergere di Cray Blitz, un programma ideato per i supercalcolatori Cray, e di Hitech, una macchina dedicata che calcolava le successioni di mosse per mezzo di 64 microelaboratori (uno per ciascuna casella della scacchiera). Hitech vinse nel 1985 il Campionato di scacchi nordamericano per calcolatori, mentre l'anno successivo Cray Blitz divenne campione del mondo fra i calcolatori, battendo proprio Hitech nella «bella» per il titolo. Cray Blitz e Hitech erano in grado di esaminare rispettivamente 100 000 e 120 000 posizioni al secondo.

La storia di Deep Thought è piuttosto singolare. In primo luogo, questa macchina è stata sviluppata da un gruppo di studenti specializzandi senza alcun appoggio ufficiale o supervisione diretta di membri della Carnegie-Mellon University. (Il gruppo di ricerca che lavora sui calcolatori che giocano a scacchi presso questa università non ha alcuna connessione con gli ideatori di Deep Thought.) In secondo luogo, i membri del gruppo non hanno tutti la stessa formazione e probabilmente proprio questa circostanza ha consentito loro di affrontare il problema in maniera originale.

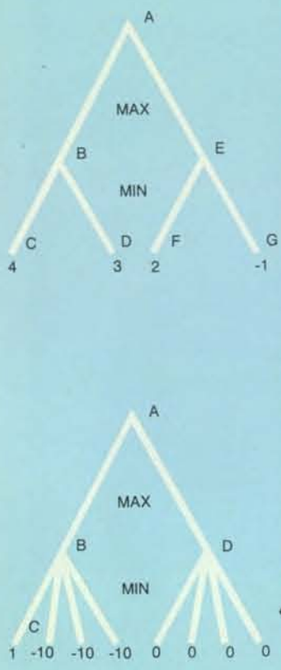
Nel giugno 1985 uno di noi (Hsu) scoprì che sarebbe stato possibile costruire un generatore di mosse su un singolo chip sfruttando le tecnologie di integrazione su grandissima scala (VLSI) fornite alla comunità scientifica dalla Defense Advanced Research Programs Agency (DARPA). Hsu trasse ispirazione per il suo chip dal generatore di mosse di Belle, ma vi introdusse diversi perfezionamenti per rendere il progetto compatibile con le tecnologie VLSI. Il chip venne concepito in modo che i componenti elettronici (fra cui 35 925 transistori) potessero essere sistemati sulla superficie più piccola possibile, nonostante la mediocre qualità del silicio fornito dalla MOSIS, la società che offriva i servizi di fabbricazione per i progetti sovvenzionati dalla DARPA (la dimensione minima di un componente era di tre micrometri). Hsu impiegò sei mesi per il lavoro di progettazione, verifica e messa a punto, ma poi dovette aspettare altri quattro mesi prima di ricevere i primi esemplari funzionanti. Il collaudo del chip fu eseguito collegandolo a una stazione di lavoro scientifica; riuscimmo così ad appurare che esso era in grado di elaborare più di due milioni di mosse al secondo, a una velocità oltre 10 volte maggiore della schiera di 64 microelaboratori dell'Hitech.

A questo punto Hsu iniziò a collaborare con Thomas Anantharaman, che all'epoca si stava specializzando in scienza dei calcolatori presso il gruppo di ricerca sul riconoscimento del linguaggio della Carnegie-Mellon University. Anantharaman aveva scritto un proprio programma per gli scacchi che generava le mosse attraverso un pacchetto di software. Sostituendo quest'ultimo con il congegno messo a punto da Hsu, Anantharaman riuscì a migliorare del 500 per cento la velocità di analisi del programma, giungendo fino a 50 000 posizioni al secondo.

Hsu e Anantharaman, ringalluzziti dal successo, decisero di preparare la macchina per farla partecipare all'edizione 1986 del Campionato nordamericano di scacchi per calcolatori, che doveva avere inizio di lì a sette settimane. Murray Campbell e Andreas Nowatzyk, che si stavano specializzando in scienza dei calcolatori, vennero chiamati a far parte del gruppo. Era importante poter disporre di una funzione di valutazione più raffinata e Campbell, che aveva esperienza di gioco agonistico, accettò di farsene carico. Il secondo e più impegnativo compito, data la limitatezza del tempo disponibile, era quello di potenziare il congegno in modo che potesse fungere da semplice strumento di ricerca, al fine di sfruttare più a fondo di quanto non fosse possibile con la stazione di lavoro la velocità potenziale del chip generatore di mosse.

Hsu dovette adottare misure draconiane per poter ultimare il progetto in tempo utile, tanto da decidere che la macchina avrebbe ignorato due regole

L'ANALISI DELL'ALBERO DI GIOCO



Un programma scacchistico genera un albero di gioco che contiene tutte le mosse possibili a partire da una posizione data, valuta le posizioni risultanti e risale a determinare la migliore prima mossa. Alle varie posizioni vengono assegnati punteggi utili dal punto di vista del programma: un punteggio elevato corrisponde a una posizione vantaggiosa.

Una tecnica molto semplice, detta algoritmo minimax, seleziona solo le linee di gioco migliori cercando i punteggi massimi per le mosse del calcolatore e quelli minimi per le risposte dell'avversario. Nell'albero in alto, per esempio, i nodi B ed E sono minimizzati a 3 e −1 rispettivamente, mentre il nodo A è massimizzato a 3.

Con questa semplice procedura è possibile analizzare un ulteriore livello dell'albero solo aumentando la potenza di calcolo di un fattore 38 (pari al numero medio di mosse possibili in una tipica posizione sulla scacchiera). La potatura «alfa-beta» perfeziona l'algoritmo consentendo al programma di ignorare linee di gioco irrilevanti, cosicché è sufficiente un potenziamento di un fattore sei per analizzare un livello in più. Se per esempio il calcolo dei punteggi comincia nel nodo C e procede verso destra, il programma assegna a B il valore 3 e trova che E ha un valore inferiore o uguale a 2; pertanto conclude che è inutile considerare F.

Un altro metodo, detto dell'«estensione particolare» sfrutta meglio le capacità del calcolatore concentrandosi sulle posizioni critiche. Nell'albero in basso, il valore di B dipende fortemente da quello di C, mentre il valore di D non è legato a nessuno dei valori delle posizioni derivate. Per migliorare l'affidabilità del punteggio di A l'algoritmo analizza quindi un'ulteriore semimossa successiva a C. Questo metodo permette a Deep Thought di spingere l'analisi a profondità straordinaria in molte posizioni tatticamente complesse.

fondamentali degli scacchi: l'arrocco e la patta per ripetizione della posizione. (Si ha questo esito quando in una partita una stessa posizione si ripresenta per tre volte e la mossa tocca sempre al medesimo giocatore.) Per compensare almeno in parte queste omissioni adottammo una strategia di ricerca ibrida: i primi livelli dell'albero di gioco venivano esaminati da un calcolatore ausiliario, che tenesse conto dell'arrocco e della patta per ripetizione della posizione, mentre i livelli successivi (che ovviamente comprendevano la maggior parte delle posizioni) venivano analizzati dalla nostra macchina.

Data la scarsità di fondi, realizzammo la nostra prima macchina, ChipTest, utilizzando componenti e sistemi che facevano parte di altri progetti; il valore complessivo degli elementi non superava i 500 dollari (1000 tenendo conto del costo dei chip costruiti con i fondi della DARPA). Ma né il nostro congegno né il calcolatore ausiliario erano perfettamente a punto all'inizio del campionato, sicché ChipTest riuscì soltanto a pareggiare vittorie e sconfitte. Non si trattava in ogni caso di un risultato scadente, per sole sette settimane di lavoro.

Questa esperienza ci permise di imparare molte cose. Hsu osservò, per esempio, che altri due programmi partecipanti al campionato si erano impegnati in linee di gioco in cui tutte le mosse apparivano forzate e nessuna delle due parti poteva prevedere l'esito finale. In altri termini, il programma che otteneva un vantaggio lo doveva unicamente al caso. Hsu propose di correggere questo difetto introducendo quello che egli chiama algoritmo di estensione particolare: nel caso in cui il calcolatore trovi una sola risposta adeguata l'analisi viene spinta a maggiore profondità, allo scopo di studiare con particolare attenzione le posizioni critiche (*si veda la «finestra» in questa pagina*).

Quando uno dei giocatori sta per catturare un alfiere intrappolato, per esempio, il difensore ha a disposizione un numero sempre minore di mosse corrette via via che la profondità di analisi aumenta; alla fine vi è una sola risposta possibile, dopo di che l'alfiere è perso. L'algoritmo di estensione particolare si applica a casi di questo tipo; in una partita ha permesso al calcolatore di stupire un maestro annunciando un matto in 19 mosse.

Anantharaman, l'unico fra noi capace di comprendere il programma da lui scritto per il calcolatore ausiliario di ChipTest, programmò da solo l'algoritmo di estensione particolare, mentre Hsu completava il microcodice, ossia l'insieme di istruzioni che controllano l'hardware al livello più elementare. Con una capacità di analisi di 400 000-500 000 posizioni al secondo, ChipTest vinse nettamente l'edizione 1987 del Campionato nordamericano di scacchi per calcolatori, riuscendo a sconfiggere fra gli altri anche la macchina campione del mondo, Cray Blitz. In questo modo si concluse l'era della forza bruta; oggi quasi tutti i programmi migliori comprendono almeno alcuni elementi di ricerca selettiva.

Il nostro lavoro aveva indicato che vi era un notevole margine di miglioramento per quanto riguardava la velocità e la capacità di analisi di ChipTest. La somma di circa 5000 dollari necessaria per lo sviluppo del nuovo progetto - chiamato Deep Thought - ci venne generosamente messa a disposizione dal relatore di Hsu, H. T. Kung.

La versione base di Deep Thought comprende 250 chip, fra cui due microelaboratori, montati tutti su un singolo pannello grande quanto metà di una pagina di questa rivista (*si veda l'illustrazione a pagina 18*). Il congegno è gestito da un programma - il cosiddetto software ospite - che «gira» su una stazione di lavoro. I microelaboratori non sono in sé più veloci di quelli montati su ChipTest, ma i perfezionamenti dell'algoritmo di ricerca permettono loro di operare con un 30 per cento di efficienza in più.

Il sistema di valutazione ha quattro componenti. Un circuito di valutazione della disposizione delle figure (il solo ereditato da ChipTest) considera la posizione di queste ultime rispetto al centro della scacchiera, la loro mobilità e altri fattori. Un secondo circuito valuta lo schieramento dei pedoni in base a parametri quali il sostegno vicendevole, il controllo del centro e la protezione del re. Un terzo circuito riconosce i pedoni «passati», ossia quelli che possono essere fatti avanzare fino all'ottava traversa e promossi a regina senza incontrare l'opposizione di pedoni avversari. Infine un circuito di valutazione delle colonne assegna valori a configurazioni più complesse di pedoni e torri su una data colonna.

Da allora abbiamo anche cominciato a considerare in che modo sia possibile affinare i circa 120 parametri della funzione di valutazione specificati dal software. Tradizionalmente i programmatori stabiliscono a priori, in base a considerazioni di carattere teorico, i «pesi» da assegnare ai pezzi e alle posizioni sulla scacchiera; riteniamo che Deep Thought sia il solo programma importante capace di regolare automaticamente questi parametri.

Abbiamo selezionato 900 partite disputate da Maestri, decidendo di definire ottimali quei pesi grazie ai quali si otteneva la migliore corrispondenza fra le mosse prescelte dalla macchina e quelle effettivamente giocate dai Maestri. Il programma di valutazione è stato quindi completamente riscritto da Campbell e Nowatzyk in base a questi criteri: per autoregolarsi, la funzione di valutazione invece di assegnare semplicemente un valore numerico finale a ciascuna posizione sulla scacchiera determina un'equazione contenente un certo numero di termini di primo grado (vale a dire, un vettore).

Sono stati utilizzati due meccanismi di autoregolazione. Il primo, che chiamia-

ANATOLIJ KARPOV CONTRO DEEP THOUGHT

HARVARD UNIVERSITY, FEBBRAIO 1990

Commenti del maestro internazionale Michael Valvo

BIANCO KARPOV	NERO DEEP THOUGHT
1 e2-e4	c7-c6
2 d2-d4	d7-d5
3 Cb1-d2	g7-g6
4 c2-c3	Af8-g7
5 e4-e5	

Con questa mossa Deep Thought esce dal suo repertorio di aperture.

...	f7-f6!

Una mossa eccellente per una macchina. Il nero attacca immediatamente il più avanzato dei pedoni bianchi e indebolisce il controllo delle caselle bianche da parte del suo avversario.

6 f2-f4	Cg8-h6
7 Cg1-f3	0-0
8 Af1-e2	f6xe5!

Deep Thought effettua una contromossa attiva.

9 f4xe5	c6-c5!

In questo modo la presa d4xc5 consentirebbe al nero di ottenere un vantaggio con Ch6-g4!

10 Cd2-b3	c5xd4
11 c3xd4	Cb8-c6
12 0-0	Dd8-b6
13 Rg1-h1	a7-a5

I due pedoni centrali danno ancora al bianco un piccolo vantaggio, ma il pedone d4 è debole e le figure non cooperano adeguatamente né occupano caselle vantaggiose.

14 a2-a4	Ac8-f5
15 Ac1-g5	Af5-e4
16 Cb3-c5!	

Questa mossa, che minaccia un attacco simultaneo alla torre e alla regina nere alla prossima mossa, tiene probabilmente conto di come i calcolatori sono programmati. Deep Thought non resiste alla tentazione di catturare il pedone.

...	Db6xb2?

È uno sbaglio; sarebbe stato meglio 16 ... Ch6-f5!, in preparazione al sacrificio di una torre con 17 Cc5-d7, Db6xb2; 18 Cd7xf8, Cf5xd4!; a questo punto l'attacco del nero sarebbe stato molto pericoloso.

17 Cc5xe4	d5xe4

Deep Thought ritiene ora di essere in svantaggio dell'equivalente di un terzo di pedone. In seguito Karpov ha dichiarato di essere rimasto impressionato dalla capacità di valutazione del calcolatore, che corrispondeva alla sua.

18 Ta1-b1	Db2-a3

Mossa forzata, dato che a 18 ... Db2-c3 seguirebbe 19 Tb1-b3, mentre 18 ... Db2-a2 dovrebbe fronteggiare 19 Cf3-d2 seguita da 20 Ae2-c4.

19 Ag5-c1	Da3-c3
20 Ac1-d2	Dc3-a3
21 Ad2-c1	Da3-c3
22 Tb1-b3	Dc3-a1

Karpov ripete le mosse per guadagnare tempo per riflettere.

23 Ae2-c4 (+)	Rg8-h8
24 Ac1xh6!	Da1xd1
25 Ah6xg7 (+)	Rh8xg7
26 Tf1xd1	e4xf3
27 g2xf3	

Sarebbe stato meglio 27 Tb3xb7; ma come poteva Karpov indovinare la mossa successiva di Deep Thought?

...	Ta8-a7!!

DEEP THOUGHT

KARPOV

Il pubblico deride questa mossa, che ritiene disastrosa, ma secondo Karpov è l'unica scelta possibile per Deep Thought.

28 Ac4-d5	Tf8-d8
29 Tb3-b5	Ta7-a6!

Il calcolatore si difende molto bene. Ora minaccia 30 ... Cc6-a7; 31 Ad5xb7, Ca7xb5; 32 Ab7xa6, Td8xd4, che porterebbe la situazione in parità.

30 Ad5-c4	Ta6-a7
31 Ac4-d5	Ta7-a6
32 Tb5-c5	Td8-d7
33 Rh1-g2	Ta6-b6!
34 Ad5xc6	b7xc6
35 Rg2-f2!	

Una mossa arrischiata, data la posizione leggermente superiore del nero, ma Karpov vuole ancora vincere.

...	Td7-d5
36 Tc5xd5	c6xd5
37 Td1-c1	Tb6-b4
38 Rf2-e3	Tb4xa4

Un altro modo, forse più semplice, per ottenere la patta sarebbe 38 ... Tb4-b3 (+); 39 Re3-e2, Tb3-b4 con ripetizione della posizione, dato che il bianco non può certo permettersi di perdere il pedone nella colonna d.

39 Tc1-c5	e7-e6
40 Tc5-c7 (+)	Rg7-g8
41 Tc7-e7	Ta4-a3 (+)
42 Re3-f4	Ta3-d3
43 Te7xe6	Td3xd4 (+)
44 Rf4-g5	Rg8-f7!

Una buona manovra difensiva.

45 Te6-a6	a5-a4

Deep Thought ritiene di essere in vantaggio e quindi non vuole forzare la patta con 45 ... h7-h6 (+); 46 Rg5xh6, Td4-h4 (+); 47 Rh6-g5, Th4-h5 (+); 48 Rg5-f4, Th5-f5 (+); seguita da 49 ... Tf5xe5.

46 f3-f4	h7-h6 (+)
47 Rg5-g4	Td4-c4?

Con 47 ... g6-g5! vi sarebbe stata patta, ma la macchina, ritenendosi ancora in vantaggio, non vuole perdere un pedone in cambio della sicurezza.

48 h2-h4	Tc4-d4
49 Ta6-f6 (+)	Rf7-g7
50 Tf6-a6	Rg7-f7
51 h4-h5	g6xh5?

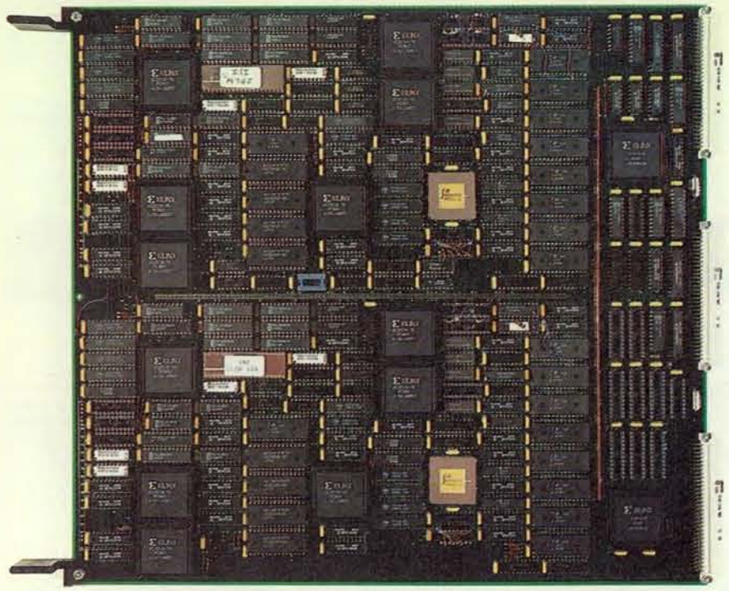
Questa mossa sancisce la sconfitta. 51 ... g6-g5, secondo Karpov, avrebbe concesso al bianco un margine di manovra troppo esiguo per vincere.

52 Rg4-f5	Rf7-g7
53 Ta6-a7 (+)	Rg7-f8
54 e5-e6	Td4-e4
55 Ta7-d7	Te4-c4
56 Td7xd5	h5-h4
57 Td5-d3	Rf8-e7
58 Td3-d7 (+)	Re7-f8
59 Td7-h7	h6-h5
60 Rf5-e5	h4-h3
61 f4-f5	Rf8-g8
62 Th7xh5	a4-a3
63 Th5xh3	a3-a2

Un giocatore preferirebbe 63 ... Tc4-a4 per prolungare la lotta, ma in questo caso 64 Re5-f6 avrebbe rapidamente la meglio.

64 Th3-a3	Tc4-c5 (+)
65 Re5-f6	Abbandona

Deep Thought si considera in svantaggio dell'equivalente di almeno sei pedoni, cosicché i suoi programmatori decidono di abbandonare. A questo punto la macchina ha ancora circa 20 minuti di tempo a disposizione, mentre a Karpov rimane meno di un minuto; tuttavia è più che sufficiente, per un giocatore del suo calibro, per conseguire la vittoria.



I principali circuiti di Deep Thought stanno in un pannello grande quanto metà di una pagina di questa rivista. Ognuno dei suoi due microelaboratori analizza fino a 500 000 posizioni al secondo. Il successore di Deep Thought, ora allo studio, sarà un sistema ad alto grado di parallelismo con 1000 chip progettati specificamente per il gioco degli scacchi.

mo meccanismo «di arrampicata», fissa arbitrariamente il valore di uno dei parametri di valutazione e, per le diverse posizioni della sua base di dati di 900 partite, effettua l'analisi di cinque o sei livelli dell'albero di gioco al fine di trovare le mosse che la macchina giocherebbe in quelle situazioni. Modifica poi leggermente il parametro e ripete il calcolo. Se il numero di corrispondenze fra le mosse prescelte dal calcolatore e quelle effettivamente giocate aumenta, il parametro viene ulteriormente modificato nello stesso senso e il processo continua fino a che tutti i parametri siano stati ottimizzati. Tuttavia per applicare questo metodo a tutti i parametri individuati occorrerebbero anni di lavoro e quindi lo abbiamo riservato solamente ai casi difficili.

Il secondo meccanismo di autoregolazione, proposto e messo in atto da Nowatzyk, è molto più rapido. Esso si basa semplicemente sulla ricerca della migliore corrispondenza fra il valore della funzione di valutazione calcolato dalla macchina e il valore presunto vero. A questo scopo si determina lo scarto minimo fra il quadrato della funzione calcolata e quello del valore vero, che per i nostri scopi è ottenuto con una ricerca spinta particolarmente in profondità (se si vuole affinare un punto particolare) o da un confronto fra le decisioni prese dalla macchina e quelle assunte da giocatori di alto livello.

Le 900 partite della base di dati forniscono ottime informazioni sul valore relativo di posizioni diverse: in definitiva, una posizione raggiunta in seguito a una mossa di un Grande Maestro è presumibilmente migliore di tutte le posizioni prodotte da mosse alternative. Anziché calcolare il valore di una posizione a partire dai parametri, Nowatzyk ha calcolato i parametri stessi in base alla differenza postulata fra la posizione prescelta da un Grande Maestro e quelle da lui respinte. L'algoritmo richiede un tempo di calcolo di pochi giorni e, contrariamente a quello «di arrampicata», non ottimizza i parametri uno per volta ma li considera simultaneamente tutti.

La nostra funzione di valutazione ad autoregolazione automatica non è né migliore né peggiore delle funzioni utilizzate dai più noti fra i programmi per gli scacchi sviluppati in ambito universitario, come Hitech e Cray Blitz. Sembra esserci ancora una notevole differenza, tuttavia, fra le capacità di analisi di Deep Thought e quelle delle migliori macchine commerciali per gli scacchi, che di solito sono il frutto di anni di lavoro da parte di esperti. Perfezionando le procedure di autoregolazione, speriamo di ridurre ben presto il divario.

Può apparire strano che la nostra macchina, pur basandosi su conoscenze scacchistiche relativamente modeste, riesca a battere giocatori eccellenti. Tuttavia occorre ricordare che il calcolatore non simula i processi di pensiero umani, ma raggiunge gli stessi scopi per vie diverse. Deep Thought vede in profondità ma osserva poco, ricorda ogni cosa ma non impara nulla, non fa mai errori madornali ma non migliora le proprie prestazioni. Tuttavia può a volte scoprire combinazioni che neppure un Grande Maestro riesce a vedere.

Queste particolari capacità spiegano forse perché il Grande Maestro Kevin Spraggett abbia deciso di allenarsi con Deep Thought in vista dell'incontro con il Grande Maestro Artur Yusupov per il quarto di finale del campionato dl mondo. Anche se la partecipazione della macchina non ha avuto effetti rilevanti sull'esito dell'incontro, costituisce comunque un precedente interessante.

Nell'ottobre 1989 a New York una versione sperimentale di Deep Thought con sei microelaboratori giocò un incontro di esibizione in due partite contro Kasparov. Questo nuovo modello era in grado di analizzare più di due milioni di posizioni al secondo, ma Kasparov lo sconfisse senza difficoltà. Benché il risultato non fosse inatteso, il gioco di Deep Thought risultò piuttosto deludente.

Nel febbraio 1990 Deep Thought si esibì nuovamente incontrando Anatolij Karpov, ex campione del mondo e attualmente sfidante di Kasparov nella finale del campionato del mondo di scacchi. I difetti manifestatisi nel programma sperimentale delle versioni a sei e a quattro microelaboratori ci indussero a ripresentare il modello a due microelaboratori. Deep Thought, potendo beneficiare di vari perfezionamenti della funzione di valutazione, giocò una delle sue migliori partite per le prime 50 mosse, poi gettò al vento una patta quasi sicura. Una versione più stabile a sei microelaboratori avrebbe avuto una velocità sufficiente per evitare questo errore grossolano (*si veda la «finestra» a pagina 16*).

La velocità d'analisi è l'aspetto delle nostre ricerche, condotte presso il Thomas J. Watson Research Center dell'IBM, a cui attualmente dedichiamo maggiore attenzione; il nostro obiettivo è di migliorare almeno di un fattore 1000 la versione attuale di Deep Thought. Il calcolatore al quale stiamo lavorando esaminerà quindi più di un miliardo di posizioni al secondo e potrà spingere la ricerca a una profondità di 14-15 livelli dell'albero di gioco nella maggior parte dei casi e addirittura a 30-60 livelli nelle successioni di mosse forzate. Se continuerà a valere la relazione osservata fra velocità di analisi e livello di gioco, la macchina della nuova generazione dovrebbe conseguire una valutazione di circa 3400 punti, ossia 800 punti al di sopra del suo livello attuale e ben 500 punti al di sopra del livello di Kasparov.

A questo fine Hsu sta progettando un chip specifico per il gioco degli scacchi che dovrebbe calcolare almeno tre milioni di mosse al secondo, con una velocità più di tre volte superiore a quella attuale di Deep Thought; è inoltre impegnato nell'ideazione di un sistema ad alto grado di parallelismo che sarà composto da 1000 di questi chip, con un ulteriore guadagno di circa 300 volte. Anantharaman e Campbell stanno perfezionando diversi aspetti della versione attuale di Deep Thought, in vista di una loro introduzione nel futuro calcolatore, e Nowatzyk persegue altri interessi di ricerca.

Noi riteniamo che questo nuovo sistema sarà sufficientemente forte, grazie alla sua sola velocità, da costituire un valido avversario per il campione del mondo; e siamo convinti che l'adozione di una lunga lista di miglioramenti già previsti consentirà alla macchina di avere la meglio, forse già nel 1992.

Kasparov è di avviso contrario e noi rispettiamo la sua opinione. In privato egli ha ammesso che una macchina in grado di analizzare un miliardo di posizioni al secondo potrebbe sconfiggere un Grande Maestro di medio livello, «ma non Karpov e me!». Il campione del mondo afferma che un giocatore di altissimo livello dovrebbe riuscire a prepararsi adeguatamente per trarre vantaggio dalle particolari carenze che sono caratteristiche dei calcolatori; a parere di Kasparov la creatività e l'immaginazione umana, e in particolare la *sua* creatività e immaginazione, non potranno che avere la meglio su un semplice congegno a base di silicio.

Di fatto, quando le due parti si troveranno di fronte sulla scacchiera, l'ingegno di un giocatore di grandissimo talento sarà opposto non a una macchina, ma al lavoro di generazioni di matematici, informatici e tecnici. Riteniamo che l'esito dello scontro ci rivelerà non se le macchine siano in grado di pensare, ma piuttosto se uno sforzo umano collettivo possa avere la meglio anche sugli individui più dotati.

BIBLIOGRAFIA

ZOBRIST ALBERT L. e CARLSON FREDERIC R., Jr., *Il calcolatore a lezione di scacchi*, in «Le Scienze» n. 62, ottobre 1973.

CONDON H. J. e THOMPSON KEN, *Belle* in *Chess Skill in Man and Machine* (seconda edizione), a cura di P. W. Frey, Springer-Verlag, 1984.

SLATE DAVID J. e ATKIN LAWRENCE R., *Chess 4.5 - The Northwestern University Chess Program* in *Chess Skill in Man and Machine* (seconda edizione), a cura di P. W. Frey, Springer-Verlag, 1984.

HSU FENG-HSIUNG, *Large Scale Parallelization of Alpha-Beta Search: An Algorithmic and Architectural Study with Computer Chess*, Carnegie-Mellon University Computer Science Department, CMU-CS-90-108, febbraio 1990.

I GRANDI SCIENZIATI

Ai personaggi che hanno fatto fare passi da gigante al pensiero umano e ai retroscena delle loro scoperte

LE SCIENZE

edizione italiana di SCIENTIFIC AMERICAN

ha dedicato numerosi articoli tra cui:

COPERNICO E TYCHO BRAHE
di O. Gingerich
(n. 67, marzo 1974)
La scoperta della copia del libro di Copernico annotata da Tycho Brahe.

ALFRED WEGENER E L'IPOTESI DELLA DERIVA DEI CONTINENTI
di A. Hallam
(n. 82, giugno 1975)
Nel 1912 questo scienziato avanzava l'ipotesi che i continenti si muovono e proponeva una teoria della loro migrazione.

GAUSS
di I. Stewart
(n. 111, novembre 1977)
Si trovava a suo agio sia con le astrazioni della teoria dei numeri sia con i lunghi calcoli dell'astronomia e le applicazioni pratiche della fisica.

LA MELA DI NEWTON E IL DIALOGO DI GALILEO
di S. Drake
(n. 146, ottobre 1980)
Fu probabilmente un diagramma visto nei *Massimi Sistemi* di Galileo a far sì che Newton collegasse la caduta della famosa mela al moto orbitale della Luna.

SADI CARNOT
di S.S. Wilson
(n. 158, ottobre 1981)
È noto per l'analisi di una macchina termica ideale, ma i suoi interessi erano rivolti alle applicazioni pratiche dell'energia del vapore.

LA BREVE VITA DI ÉVARISTE GALOIS
di T. Rothman
(n. 166, giugno 1982)
Secondo la leggenda il giovane matematico formulò di getto la teoria dei gruppi nella notte precedente il duello che lo portò alla morte.

GEORG CANTOR E LA TEORIA DEGLI INSIEMI TRANSFINITI
di J.W. Dauben
(n. 180, agosto 1983)
Cantor ha dimostrato che esiste una gerarchia di infiniti, ciascuno più «grande» del precedente; oggi la sua teoria degli insiemi è uno dei cardini della matematica.

MONDINO DE' LIUZZI
di P.L. Lollini e L. Pelegatti
(n. 182, ottobre 1983)
Ebbe il merito di rinnovare la scienza medica medievale introducendo nell'aula universitaria la dissezione a scopo didattico e di ricerca.

DARWIN GEOLOGO
di S. Herbert
(n. 215, luglio 1986)
Nei cinque anni passati sulla *Beagle* la sua principale attività di ricerca riguardò la geologia.

WILLIAM HERSCHEL E LA NASCITA DELL'ASTRONOMIA MODERNA
di M. Hoskin
(n. 216, agosto 1986)
Mediante i telescopi da lui stesso costruiti Herschel scoprì migliaia di stelle e nebulose.

I CONTRIBUTI DI LEONARDO ALLA MECCANICA TEORICA
di V. Foley e W. Soedel
(n. 219, novembre 1986)
Un esame approfondito dei *Codici di Madrid* fa emergere l'importanza delle intuizioni di Leonardo su quattro aspetti fondamentali della meccanica.

ANDRÉ-MARIE AMPÈRE
di L. Pearce Williams
(n. 247, marzo 1989)
Fu il primo ricercatore a valutare quantitativamente gli effetti magnetici della corrente elettrica, ma si interessò anche di filosofia della scienza.

Imprinting parentale nei geni

Anche quando padre e madre trasmettono ai figli geni identici, questi possono produrre effetti diversi a seconda del genitore da cui provengono; questo fenomeno può influire sullo sviluppo e innescare varie patologie

di Carmen Sapienza

L'osservazione che Gregorio Mendel fece alla metà del XIX secolo, nel corso dei suoi primi esperimenti sui piselli, sarebbe diventata in seguito un assioma per i genetisti. Mendel si era accorto che, incrociando piante di pisello «pure» con semi lisci e piante di pisello «pure» con semi grinzosi, si ottenevano piante «ibride» equivalenti, ma tutte con semi lisci. Il risultato era il medesimo sia che la pianta con piselli lisci rappresentasse la generazione parentale maschile sia, inversamente, che rappresentasse la generazione parentale femminile. Dunque, Mendel aveva scoperto il principio dell'equivalenza negli incroci reciproci: il gene, indipendentemente dal genitore che lo trasmette alle prole, si comporta sempre allo stesso modo.

Nella storia e nella pratica della genetica l'importanza dell'osservazione di Mendel non può certo essere sottovalutata. Essa è infatti risultata valida per un gran numero di caratteri genetici, riguardanti non soltanto i piselli, ma anche i moscerini della frutta, i topi, gli esseri umani e tutta una schiera di altri organismi.

Di recente, tuttavia, alcuni genetisti ed embriologi hanno descritto un certo numero di caratteri che non seguono questa norma: tali caratteri sono infatti controllati da un fenomeno cui è stato dato il nome di «imprinting genomico». Si tratta di un processo che, in maniera temporanea e reversibile, lascia un'impronta di diverso tipo nei geni trasmessi dal genitore di sesso maschile e in quelli trasmessi dal genitore di sesso femminile. Di conseguenza, la prole che riceve geni marcati dalla madre sarà diversa da quella che riceve geni marcati dal padre. In poche parole, è in qualche caso importante il fatto che un gene venga trasmesso da un genitore piuttosto che dall'altro.

Ricercatori di un gran numero di laboratori sparsi un po' in tutto il mondo stanno cercando da tempo di identificare la natura molecolare di questo imprinting che si instaura nel genoma, nonché il meccanismo attraverso cui si realizza, e il numero e i tipi di geni che esso contrassegna. Il lavoro del nostro gruppo di ricerca, benché non sia stato ancora coronato da successo, ci ha consentito già di fare alcune affascinanti scoperte, che hanno arricchito le nostre conoscenze su determinate forme di cancro, su alcune malattie genetiche e su altre patologie. Esso inoltre ha messo in luce insospettate influenze sull'ereditarietà di alcuni caratteri sinora considerati in base alla genetica classica mendeliana.

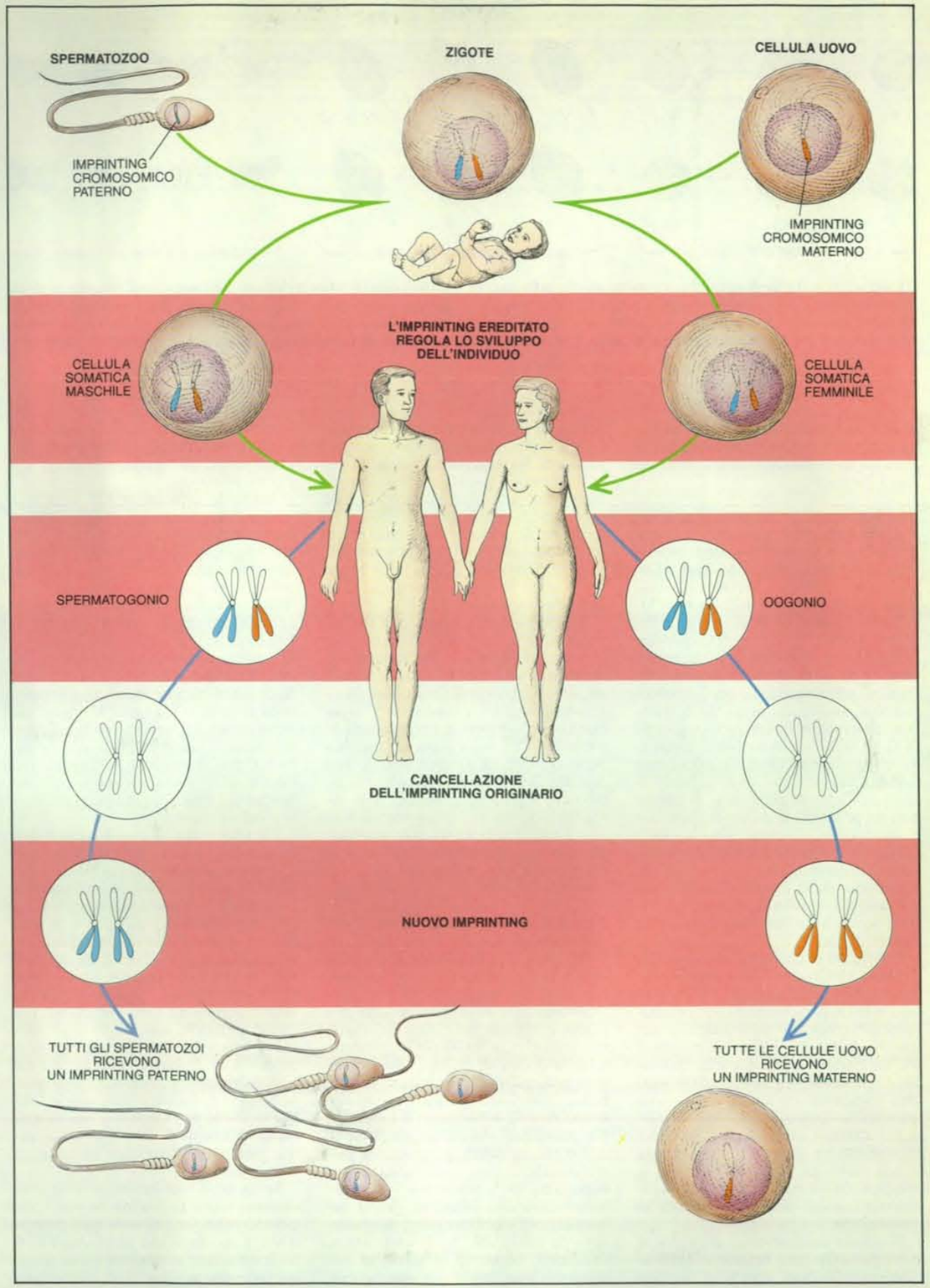
In realtà, sono note da molti anni alcune eccezioni al principio dell'equivalenza degli ibridi reciproci, ma in generale esse sono sempre state fatte rientrare nell'una o nell'altra di due categorie. Della prima categoria fanno parte i caratteri legati a geni localizzati nei cromosomi sessuali, X e Y. Le femmine dei mammiferi hanno due cromosomi sessuali X in tutti i nuclei cellulari, mentre nei maschi questi cromosomi sono diversi e sono noti come X e Y. Tra i numerosi caratteri determinati da geni presenti sul cromosoma X possiamo ricordare il daltonismo e l'emofilia. La loro trasmissione ereditaria segue uno schema ben definito, che non deve essere necessariamente equivalente negli ibridi reciproci.

Per esempio, se un uomo daltonico ha figli da una donna non portatrice di daltonismo, nessuno dei figli maschi sarà daltonico. Se la madre è daltonica e il padre no, i figli maschi saranno invece tutti daltonici. In entrambi i casi, le figlie saranno portatrici del gene per il daltonismo, ma non saranno daltoniche. La trasmissione ereditaria e la manifestazione del carattere legato al sesso dipende dal sesso della prole, ma non direttamente dal sesso del genitore che presenta quel carattere.

La seconda categoria di incroci reciproci non equivalenti contempla caratteri controllati da geni situati al di fuori del nucleo cellulare. Alcuni organelli subcellulari (i mitocondri nelle cellule animali, e i mitocondri e i cloroplasti nelle cellule vegetali) possiedono una propria informazione genetica; poiché vengono trasmessi di generazione in generazione col citoplasma della cellula uovo i loro geni vengono ereditati esclusivamente per via materna. Il colore delle foglie in alcune piante mostra questo tipo di eredità, così come, negli esseri umani, avviene per l'encefalomiopatia mitocondriale, una forma di malattia neuromuscolare.

L'imprinting genomico, il terzo tipo di eccezione al principio dell'equivalenza degli incroci reciproci (e quello scoperto più di recente), differisce nettamente dagli esempi che abbiamo riportato in precedenza. I caratteri che esso influenza non sono necessariamente determinati da geni localizzati sui cromosomi sessuali (anche se possono esserlo) e neppure sono associati a organelli ereditati per via materna. Almeno in teoria, l'imprinting può influenzare qualsiasi gene presente in una cellula. Le eccezioni legate al sesso e agli organelli materni si basano su un ineguale contributo genetico da parte di ciascun genitore. Per contro, a causa dell'imprinting genomico, anche se i genitori trasmettono alla prole geni assolutamente identici, gli effetti di questi geni non saranno uguali

I geni presenti nei cromosomi degli spermatozoi e delle cellule uovo possono aver subito un imprinting, per cui individui di entrambi i sessi possono ereditare alcuni geni «marcati» dal padre e alcuni geni «marcati» dalla madre. Tuttavia, quando questi individui producono a loro volta spermatozoi o cellule uovo, il vecchio imprinting si cancella e un nuovo imprinting specifico per il sesso dell'individuo in questione influenza tutti i cromosomi dei gameti.



INCROCI RECIPROCI

Gli incroci in cui il gene di un carattere viene trasmesso dalla madre in un caso e dal padre nell'altro sono detti reciproci. Per la maggior parte dei caratteri, come il «liscio» nei piselli, non importa quale sia il genitore che trasmette il gene (*a sinistra*). Certi caratteri, invece, come il colore delle foglie in alcune piante, vengono ereditati esclusivamente dalla madre (*al centro*). Per i caratteri legati ai cromosomi sessuali, come il colore delle ali nel lepidottero *Abraxas*, lo schema di trasmissione ereditaria è più complesso (*a destra*).

se sarà stato esercitato su di essi un imprinting diverso. L'attenzione dei biologi è stata notevolmente attratta da questi nuovi caratteri.

In buona parte, l'interesse attuale attorno ai caratteri influenzati da un imprinting genomico deriva dagli esperimenti eseguiti da James McGrath e Davor Solter al Wistar Institute of Anatomy and Biology di Filadelfia e da M. Azim H. Surani e collaboratori all'Institute of Animal Physiology di Cambridge in Inghilterra. McGrath e Solter hanno messo a punto un'elegante tecnica «chirurgica» estremamente fine, il trasferimento di nuclei, che permette di scambiare fisicamente l'informazione genetica di un embrione di topo con quella di un altro embrione. Questa tecnica è resa possibile dal fatto che, dopo la fecondazione dell'uovo da parte dello spermatozoo, ma prima che le cellule dell'embrione comincino a dividersi, il nucleo della cellula uovo e quello dello spermatozoo rimangono separati per breve tempo nel citoplasma dello zigote. Ciascuno dei due nuclei è visibile al microscopio ottico e ha una posizione e una dimensione caratteristiche.

Per mezzo di una sottilissima pipetta di vetro, McGrath e Solter sono riusciti ad asportare selettivamente dallo zigote o il nucleo appartenente all'uovo o quello appartenente allo spermatozoo (o entrambi). Sono poi riusciti a sostituire il nucleo rimosso con un nucleo proveniente da un altro zigote e hanno trovato che, sostituendo un nucleo derivato da uno spermatozoo con uno avente la stessa provenienza o facendo la stessa operazione con un nucleo ottenuto da una cellula uovo, gli embrioni risultanti si sviluppavano in modo completo ed erano indistinguibili dai topi normali dello stesso ceppo.

Mediante trasferimento di nuclei, è stato possibile dare origine a embrioni con costituzioni genetiche insolite: per esempio, embrioni con entrambe le serie di cromosomi di un unico genitore. McGrath, Solter e Surani, volendo sapere in che modo questi embrioni si sarebbero sviluppati, hanno prodotto alcuni «ginogenoti» (embrioni con due serie complete di cromosomi materni) e «androgenoti» (embrioni con due serie complete di cromosomi paterni) e ne hanno confrontato lo sviluppo con quello di embrioni normali.

È importante notare che, essendo stati incrociati tra loro da molte generazioni, i topi utilizzati in questi esperimenti avevano, sia nei maschi, sia nelle femmine, serie identiche di cromosomi (tranne, ovviamente, il cromosoma Y, mancante nelle femmine, e il cromosoma X, presente in un'unica copia nei maschi). Se lo sviluppo di un embrione di topo dipende esclusivamente dal corredo genico che l'animale possiede, non dovrebbe in teoria importare che un topo riceva tutti i geni da un solo genitore: ginogenoti, androgenoti e topi normali dovrebbero tutti svilupparsi in modo identico.

In realtà, invece, questi embrioni non si sviluppano nello stesso modo. Lo sviluppo sia degli individui con due serie di geni esclusivamente di origine materna, sia di quelli con due serie di geni esclusivamente di origine paterna non giunge a termine: generalmente si arresta dopo che la divisione ha prodotto qualche decina di cellule. Quando riesce a proseguire, come capita di tanto in tanto, i ginogenoti e gli androgenoti mostrano diverse e interessanti anomalie. Nei ginogenoti che hanno raggiunto lo stadio più avanzato, gli embrioni presentano aberrazioni relativamente poco importanti, ma la placenta e il sacco del tuorlo (strutture essenziali per il nutrimento dell'embrione) sono assai scarsamente sviluppati. Negli androgenoti che hanno raggiunto lo stadio di sviluppo più avanzato accade esattamente il contrario: il sacco del tuorlo e la placenta sembrano più vicini alla normalità, mentre gli embrioni sono gracili e poco sviluppati.

Poiché le sequenze di DNA nei cromosomi dei ginogenoti, degli androgenoti e degli embrioni normali sono sempre uguali, McGrath, Solter e Surani hanno formulato l'ipotesi che i geni fossero stati in qualche modo modificati o marcati in maniera diversa in conseguenza della loro origine paterna o materna. Era come se il procedimento di imprinting avesse disattivato selettivamente alcuni geni che sarebbero stati, invece, funzionanti durante lo sviluppo precoce degli embrioni. Tra i geni di origine paterna alcuni, importanti per lo sviluppo embrionale, sembravano inattivi, mentre tra i geni di origine materna parevano inattivi quelli aventi un'importanza critica per la formazione del sacco del tuorlo e della placenta.

Un'altra brillante dimostrazione dell'imprinting è stata data da Bruce M. Cattanach, un genetista del Medical Research Council di Oxford, in Inghilterra, il quale da anni si interessava al fenomeno della non equivalenza di certi incroci reciproci. Per studiare questi caratteri, Cattanach si servì di ceppi di topi che presentavano una o più coppie di cromosomi fusi: topi, cioè, in cui alcuni cromosomi sono fisicamente saldati e, pertanto, non hanno la possibilità di separarsi nel corso della meiosi, il processo di divisione cellulare da cui hanno origine le cellule uovo e gli spermatozoi. Di conseguenza, una cellula che si divide potrà ricevere occasionalmente due copie di uno stesso cromosoma, mentre un'altra cellula non ne riceverà alcuna.

Se un uovo con due copie di un cromosoma viene fecondato da uno spermatozoo che non possiede quel cromosoma, l'embrione che ne risulterà conterrà il numero normale di cromosomi, ma entrambe le copie di un cromosoma proverranno dalla madre. Analogamente entrambe le copie saranno di origine paterna se lo spermatozoo conterrà i due cromosomi e la cellula uovo non ne conterrà alcuno.

Secondo il principio dell'equivalenza negli ibridi reciproci, tutti gli animali che derivano da tali incroci dovrebbero essere identici. Ma Cattanach ha inequivocabilmente dimostrato, con i suoi esperimenti, che tale principio non vale per diversi cromosomi di topo.

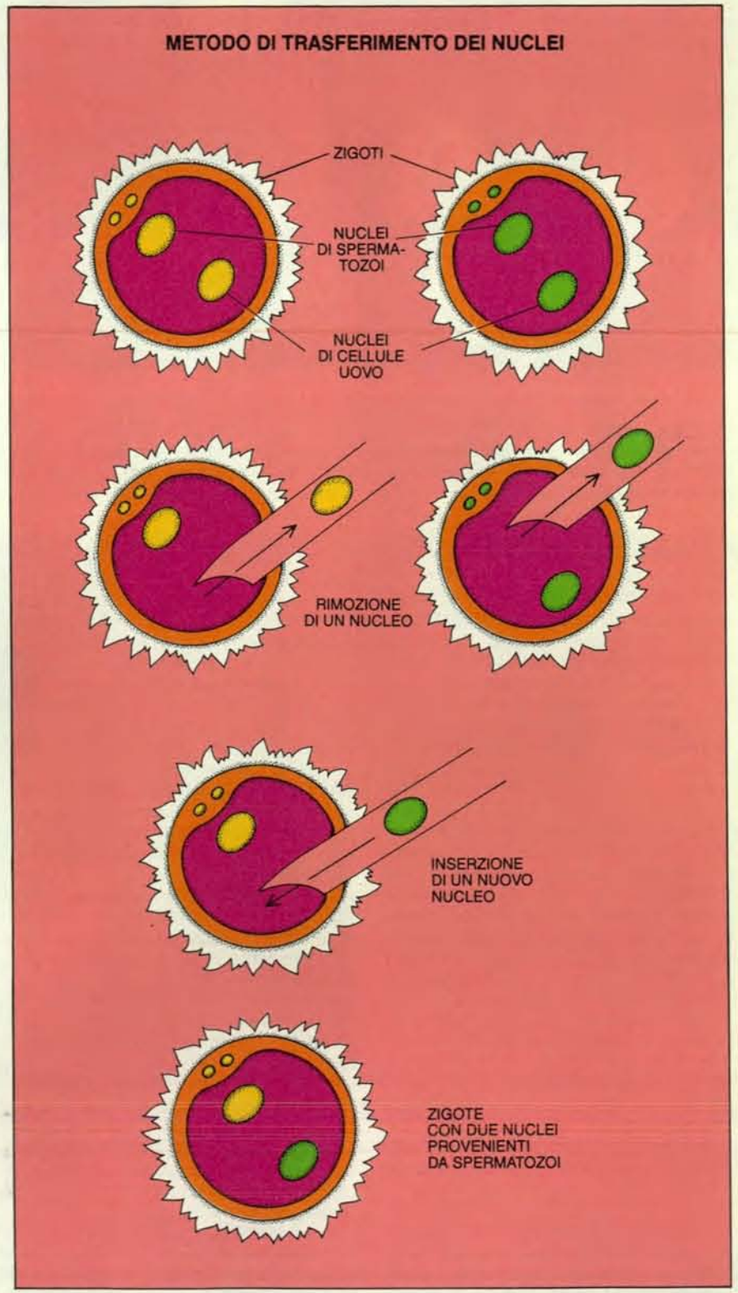
Per esempio, egli ha trovato che i topi che avevano ricevuto entrambe le copie del cromosoma 11 da un unico genitore erano radicalmente diversi da quelli normali, per la taglia e per il peso. Allevati nelle medesime condizioni, i figli con due cromosomi 11 di origine materna avevano dimensioni anormalmente piccole, mentre quelli che avevano ricevuto quei due cromosomi dal padre avevano dimensioni gigantesche. Esperimenti analoghi, effettuati con altre coppie di cromosomi fusi, hanno prodotto aberrazioni caratteristiche, che hanno portato persino alla morte dell'embrione e alla manifestazione di anomalie comportamentali.

Gli esperimenti di Cattanach sono valsi a dimostrare un altro punto importante: gli effetti dell'imprinting genomico non persistono nella generazione successiva, il che significa che il fenomeno non provoca una modificazione permanente del cromosoma. Un topo di piccole dimensioni e di sesso maschile, con entrambe le copie del cromosoma 11 provenienti dalla madre, dà vita, di norma, a prole di dimensioni normali. In un modo o nell'altro, dunque, i geni ricevuti dalla madre devono essere privati dell'impronta femminile e devono venire nuovamente contrassegnati come geni maschili.

Ispirati dalle scoperte di McGrath, Solter, Surani e Cattanach, ricercatori di parecchi laboratori (compreso il mio) hanno cercato di vedere se l'imprinting nel genoma potesse essere il risultato di una modificazione diretta del DNA. In particolare, abbiamo studiato il ruolo della metilazione del DNA, un processo in cui piccoli gruppi molecolari sono legati chimicamente alla macromolecola di DNA (si veda l'articolo *Un tipo diverso di trasmissione ereditaria* di Robin Holliday in «Le Scienze» n. 252, agosto 1989). Un fatto che ci ha molto incuriosito è stato che geni particolari vengono metilati in modo diverso a seconda che provengano dalla madre o dal padre e che le modificazioni in essi prodotte siano reversibili. Nessuno, però, è stato ancora in grado di dimostrare se queste modificazioni siano il meccanismo primario dell'imprinting genomico oppure se siano semplicemente il riflesso di qualche fenomeno biologico nascosto più importante.

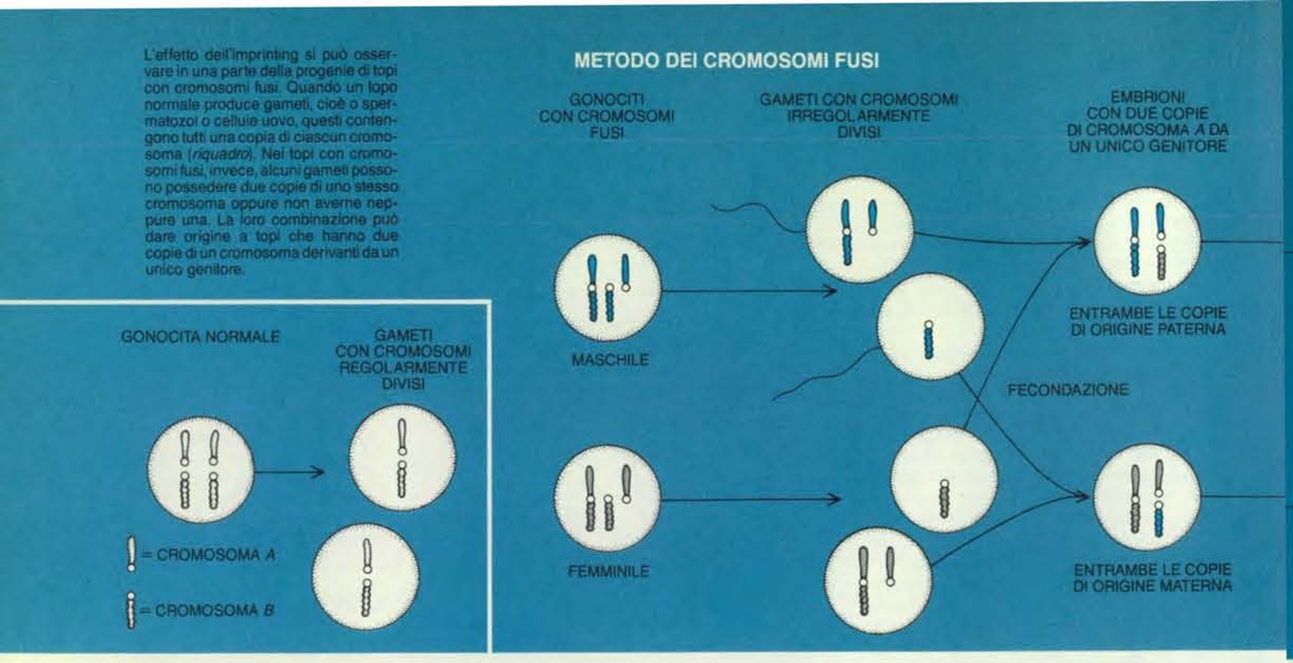
Anche se i miei colleghi non sono tutti d'accordo con me, io sono convinto che la metilazione del DNA sia solo una conseguenza del meccanismo primario. Pure senza conoscere le cause che sono alla base dell'imprinting genomico è comunque possibile effettuare molti importanti esperimenti. Purtroppo, infatti, indipendentemente dai meccanismi che ne sono all'origine, l'imprinting si manifesta in modi spesso strani e talvolta penosi.

Queste manifestazioni fanno la loro



Una tecnica per lo studio dell'imprinting genomico si basa sul metodo di trasferimento dei nuclei. Ne risultano embrioni che posseggono cromosomi provenienti da un unico genitore.

L'effetto dell'imprinting si può osservare in una parte della progenie di topi con cromosomi fusi. Quando un topo normale produce gameti, cioè o spermatozoi o cellule uovo, questi contengono tutti una copia di ciascun cromosoma (*riquadro*). Nei topi con cromosomi fusi, invece, alcuni gameti possono possedere due copie di uno stesso cromosoma oppure non averne neppure una. La loro combinazione può dare origine a topi che hanno due copie di un cromosoma derivanti da un unico genitore.



comparsa non solo negli animali da esperimento, ma anche negli esseri umani, talvolta sotto forma di alcune gravi malattie. Di recente, Robert D. Nicholls e i suoi collaboratori alla Harvard Medical School hanno effettuato una simile scoperta in pazienti affetti dalla sindrome di Prader-Willi, una malattia caratterizzata da ritardo mentale, obesità estrema, bassa statura e dimensioni sproporzionatamente ridotte delle mani e dei piedi. Dopo aver esaminato i genomi di questi pazienti e quelli dei loro genitori, Nicholls ha concluso che molti pazienti avevano ereditato entrambe le copie del cromosoma 15 dalla propria madre.

Nicholls, Joan H. M. Knoll della Harvard Medical School e Charles A. Williams dell'Università della Florida hanno effettuato una scoperta affine, ma contrapposta, in pazienti affetti da sindrome di Angelman. (In passato, questi soggetti erano descritti come marionette sempre allegre, che ridevano in modo eccessivo, si muovevano convulsamente e manifestavano altri sintomi di ritardo motorio e mentale.) I pazienti affetti da questa sindrome subiscono frequentemente parziali delezioni del cromosoma 15 di origine materna, con il risultato che solo il cromosoma 15 paterno è presente per intero.

È sorprendente che due malattie con quadri clinici così diversi possano essere collegate a un imprinting differenziale degli stessi geni situati sullo stesso cromosoma. Ma, contrariamente ai topi anormalmente grossi o piccoli, che Cattanach ha ottenuto attraverso i suoi esperimenti, la sindrome di Prader-Willi e quella di Angelman non si possono spiegare facilmente come facce opposte di una stessa medaglia, causate da un eccesso o da una carenza degli stessi prodotti genici. Gli studi di Nicholls e collaboratori mettono in evidenza quanto sia difficile prevedere in che modo caratteri particolari possano reagire all'imprinting. Pertanto può essere utile riesaminare molte malattie umane ereditarie per cercare di mettere in luce gli effetti di questo processo.

Possono venire colpiti anche caratteri per i quali non esiste alcuna ovvia necessità di chiamare in causa l'influenza dell'imprinting genomico. Questo assunto può essere illustrato da diverse forme di cancro infantile, tra cui il rabdomiosarcoma dell'embrione (un tumore che colpisce i muscoli), il tumore di Wilms (una forma di carcinoma renale) e l'osteosarcoma (una forma di cancro delle ossa).

Per spiegare come l'imprinting genomico possa trovarsi coinvolto in queste malattie, sarebbe opportuno illustrare brevemente come si pensa che compaiano le suddette forme di tumore. Molte forme di cancro sono ritenute la conseguenza di un accumulo di successive mutazioni in geni specifici all'interno di una singola cellula. Per esempio, sul cromosoma 11 si trova un gene, *Rd*, che fa parte di una famiglia di geni variamente conosciuti come oncogèni recessivi, o geni che sopprimono i tumori o anti-oncogèni (si veda l'articolo *La ricerca degli anti-oncogèni* di Robert A. Weinberg in «Le Scienze» n. 243, novembre 1988). Senza il prodotto di *Rd* una cellula muscolare si trasforma in una cellula neoplastica che poi evolve in un rabdomiosarcoma.

Dato che ogni nucleo cellulare contiene due copie del cromosoma 11, prima che la cellula si trasformi devono essere disattivate entrambe le copie di *Rd*. La prima copia può essere disattivata in molteplici modi, il più semplice dei quali è una mutazione che altera la sequenza di DNA del gene stesso. Se questa mutazione ha luogo, allora qualunque altro evento genetico che inattivi la rimanente copia di *Rd* farà probabilmente diventare cancerosa la cellula. Spesso l'evento genetico che distrugge la seconda copia di *Rd* provoca o una delezione completa del cromosoma o una delezione parziale proprio di quella parte che contiene il gene.

Con Heidi J. Scrable e Webster K. Cavenee del Ludwig Institute for Cancer Research di Montreal ho proposto, per quanto riguarda alcuni tumori maligni che colpiscono l'infanzia, un meccanismo alternativo. Secondo la nostra opinione, l'evento che inattiva la prima copia di un oncogène recessivo non sarebbe una vera e propria mutazione, ma al contrario potrebbe essere un imprinting genomico, in grado di disattivare i geni in maniera diversa nei maschi e nelle femmine. (Il secondo avvenimento, che indurrebbe la trasformazione definitiva della cellula, sarebbe probabilmente ancora una delezione cromosomica o un incidente di altra natura.) In base alla nostra ipotesi, dunque, le cellule del rabdomiosarcoma dovrebbero possedere sul cromosoma 11 sopravvissuto una copia marcata, inattiva, del gene *Rd*. Inoltre, nella maggior parte dei casi di rabdomiosarcoma embrionale questo cromosoma proverrebbe sempre dallo stesso genitore.

Nel nostro laboratorio e in quelli di Grady F. Saunders dell'M. D. Anderson Cancer Center di Houston, di Manfred Mannens dell'Università di Amsterdam, di Masao Sasaki dell'Università di Kyoto e di molti altri sono state in effetti trovate alcune prove che convalidano la nostra teoria. La prima inattivazione nel rabdomiosarcoma embrionale, nel tumore di Wilms e nell'osteosarcoma ha luogo quasi sempre in un oncogène recessivo che si trova su un cromosoma paterno. I dati ottenuti depongono in favore di un ruolo esercitato dall'imprinting genomico nella genesi di questi tumori, ma prospettano anche un problema teorico importante.

Il problema nasce dal fatto che la maggior parte di coloro, compreso il sottoscritto, che hanno studiato l'imprinting genomico, l'ha considerato come una conseguenza della notevole differenza che esiste, tra maschi e femmine, nella fisiologia e nella biochimica della formazione dei gameti (uova e spermatozoi). Nel corso della maggior parte della loro vita da adulti, i maschi continuano a produrre miliardi di spermatozoi a partire da una popolazione di cellule staminali che si dividono ininterrottamente. Per contro, le femmine possiedono già al momento della nascita tutte le cellule uovo che via via andranno maturando nel corso del periodo riproduttivo.

Se l'imprinting differenziale riflettesse soltanto la dicotomia tra produzione maschile e produzione femminile dei gameti, tutti i maschi dovrebbero marcare i loro genomi in un certo modo e le femmine in un modo diverso. Non potendo provare il contrario, abbiamo dato per scontato che tutti gli appartenenti a uno dei due sessi marcassero i rispettivi genomi in maniera identica.

Ma i dati che collegano i tumori maligni dell'infanzia a oncogèni recessivi inattivi di origine paterna hanno smentito questo nostro assunto. Per esempio, se tutti i maschi marcassero e inattivassero il gene *Rd*, l'incidenza del rabdomiosarcoma sarebbe molto elevata: una sola alterazione genetica nel gene *Rd* ereditato dalla madre farebbe trasformare alcune cellule di un individuo qualsiasi in cellule tumorigene. Poiché i tumori di questo tipo sono rari (colpiscono approssimativamente un bambino su 20 000), è improbabile che si ereditino un oncogène recessivo inattivo dal padre. Nei soggetti affetti da questo genere di malattie sembra invece che accada esattamente questo.

Una affascinante spiegazione di questa differenza è che non tutti i maschi (o, presumibilmente, tutte le femmine) marchino e inattivino gli stessi geni. Quale potrebbe essere la causa di tali differenze tra individui?

Se uno è un genetista convinto, si rivolgerà sempre in primo luogo alla genetica per ricevere risposte a interrogativi del genere. Per una risposta di natura genetica devono esserci uno o più geni responsabili dell'imprinting (da non confondersi con quei geni che subiscono essi stessi l'imprinting). In altre parole, l'imprinting deve essere un processo che nasce dall'attività di svariati geni, che operano in maniera diversa nei maschi e nelle femmine. I soggetti dello stesso sesso con geni diversi che controllano l'imprinting avranno anche costellazioni diverse di geni marcati. Per esempio, la maggior parte dei maschi non disattiverà il gene *Rd*, ma lo faranno alcuni individui con una copia aberrante di un gene che controlla l'imprinting.

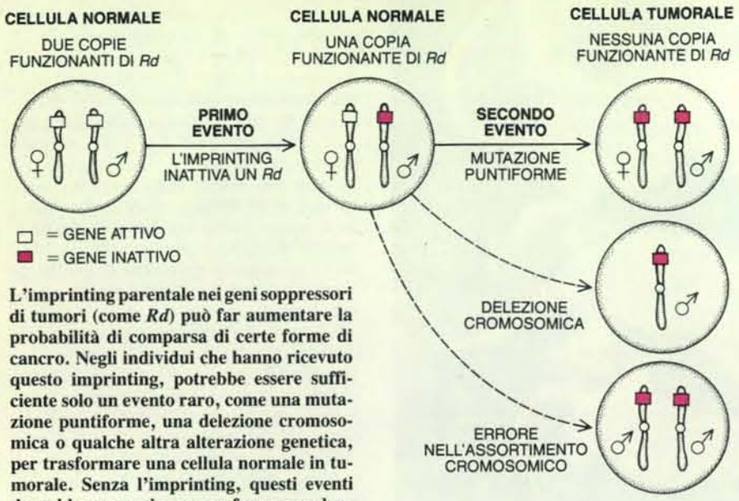
L'idea che l'espressione o la mancanza di espressione di un gene possa essere controllata da altri geni non è nuova: descrive un vecchio e ben noto fenomeno, definito modificazione della dominanza. Molti caratteri sono sensibili all'attività di altri geni che si dice ne modifichino l'espressione. L'imprinting genomico può essere visto come un caso speciale di modificazione della dominanza. L'unica caratteristica insolita dei geni modificatori che si presuppone controllino l'imprinting genomico è quella di agire in maniera diversa nei maschi e nelle femmine.

La corea di Huntington è un esempio interessante di carattere che sembra essere modificato dall'imprinting genomico sesso-specifico. Questa malattia neurologica letale viene ereditata come carattere dominante, il che significa che chiunque riceva il gene *HD* da un genitore ne rimarrà colpito. La corea di Huntington colpisce in generale gli individui di mezza età: l'età media per la diagnosi è 38 anni. All'incirca il 10 per cento dei casi, tuttavia, è caratterizzato da comparsa della malattia in età giovanile: vengono colpiti bambini di solo due anni e mezzo.

Come hanno potuto notare i ricercatori nel corso dei loro studi, circa il 90 per cento dei bambini colpiti proviene da famiglie in cui il padre è il genitore ammalato. Chiaramente, in questi casi, sia il padre, sia i figli possiedono lo stesso carattere genetico per la malattia, ma la modificazione del carattere stesso attraverso l'eredità paterna fa sì che l'età di comparsa della malattia sia molto più precoce nei figli.

Nel corso degli anni molte ipotesi sono state avanzate per spiegare il comportamento genetico della corea di Huntington, ma il modello proposto da Charles D. Laird dell'Università di Washington è quello che più si avvicina all'interpretazione dei dati. L'innovazione di Laird è consistita nel ricostruire la genetica della malattia suddetta sulla base di una classe specifica di caratteri della *Drosophila*, che vengono espressi in modo variabile: gli effetti di posizione che provocano variegature. Questi caratteri sono insoliti in quanto, diversamente dalla maggior parte degli altri caratteri mutanti, nei tessuti colpiti non vengono espressi da tutte le cellule. Questi tessuti diventano, al contrario, mosaici costituiti da alcune cellule mutate e da altre che appaiono perfettamente normali. L'aliquota delle cellule che esprimono la mutazione è regolata da geni modificatori che possono far spostare l'equilibrio da una condizione di tessuto quasi completamente normale a una condizione di tessuto completamente mutante.

Laird sosteneva che, se la variabilità nell'età di comparsa della corea di Huntington rifletteva un mosaicismo variabile nell'espressione del gene *HD*, questo mosaicismo doveva con ogni pro-



L'imprinting parentale nei geni soppressori di tumori (come *Rd*) può far aumentare la probabilità di comparsa di certe forme di cancro. Negli individui che hanno ricevuto questo imprinting, potrebbe essere sufficiente solo un evento raro, come una mutazione puntiforme, una delezione cromosomica o qualche altra alterazione genetica, per trasformare una cellula normale in tumorale. Senza l'imprinting, questi eventi dovrebbero aver luogo con frequenza doppia perché una cellula diventi neoplastica.

babilità essere influenzato da geni modificatori. Un gene modificatore che spingesse l'equilibrio del mosaicismo verso una condizione di tessuto prevalentemente mutante provocherebbe la comparsa della corea di Huntington in una età più precoce, mentre un gene modificatore che spingesse l'equilibrio del mosaicismo verso una condizione di tessuto normale porterebbe alla comparsa della malattia in età più avanzata. (Sono stati riferiti casi di pazienti affetti da corea di Huntington che non hanno manifestato la malattia prima della settantina.)

Laird ha anche spiegato che la preponderanza di casi a eredità paterna con comparsa della malattia in età giovanile è dovuta alla localizzazione del gene modificatore sul cromosoma X. Dato che i maschi hanno un solo cromosoma X, qualsiasi aberrazione del gene modificatore localizzato su di esso non verrebbe compensata come lo sarebbe nelle femmine, che hanno due cromosomi X. Quindi nei maschi c'è più probabilità che la prole abbia un mosaicismo il cui equilibrio è spostato verso una condizione di tessuto prevalentemente mutante e una età più precoce di comparsa della malattia. Il 10 per cento dei casi di corea di Huntington che compaiono in età giovanile si verifica in famiglie in cui è la madre, e non il padre, a essere colpita dalla malattia, ma questi casi sono prevedibili: dovrebbero esistere, anche se rare, femmine con geni modificatori aberranti su entrambi i cromosomi X.

Sembra che questo modello ci metta in grado di spiegare adeguatamente la genetica della corea di Huntington. Le mie ricerche hanno consentito di migliorare leggermente lo schema di Laird in modo da spiegare in maniera più dettagliata il comportamento dei geni modificatori, e questa piccola aggiunta ha offerto un avvicinamento ancora più spinto al comportamento genetico reale della corea di Huntington.

Secondo me, il modello di Laird basato sull'effetto di posizione, applicato alla genetica della corea di Huntington, riporta lo studio dell'imprinting genomico quasi al suo punto di partenza. L'effetto di posizione che causa variegature nella *Drosophila* è stato il tema del primo studio approfondito dell'imprinting genomico che io conosca: la ricerca di Janice Spofford dell'Università di Chicago. Questa ricerca, pubblicata nel 1959, ricevette scarsa attenzione. Negli ultimi trent'anni sono stati riportati sporadicamente anche altri dati su caratteri genetici che avevano ricevuto un imprinting da parte del genitore che aveva trasmesso un gene particolare. Vengono, a quanto pare, influenzati dall'imprinting genomico caratteri presenti in organismi assai diversi: moscerini della frutta, lieviti, granoturco, topi ed esseri umani.

Cionondimeno l'imprinting genomico viene considerato in genere, perfino da molti biologi, come un fatto curioso che potrebbe interessare tutt'al più un numero molto limitato di caratteri. In parecchie occasioni mi è stato chiesto perché sprecassi il mio tempo (e, pertanto, anche quello del mio interlocutore) per un fenomeno di così scarsa importanza. Ho sempre risposto che il numero di caratteri interessato dall'imprinting genomico è sconosciuto, ma probabilmente grande. Questa mia risposta generalmente provoca sbalordimento e incredulità in chi mi ascolta; segue da parte del mio interlocutore un breve predicozzo sulle leggi di Mendel culminante nell'affermazione - fatta battendo il pugno sul tavolo - che la maggior parte dei caratteri non si comporta in «quel modo».

Fino a un certo punto, queste critiche sono giuste. Se la maggior parte dei caratteri mostrasse una dipendenza assoluta dal genitore dal quale sono stati ereditati i geni responsabili di quei caratteri, i genetisti certamente lo avrebbero notato da tempo. Il punto chiave della divergenza, tuttavia, potrebbe essere il livello al quale i ricercatori esaminano un particolare carattere. Quando la Spofford studiò l'espressione a mosaico dei caratteri mutanti nei moscerini della frutta, poté osservare gli stessi caratteri in tutti i discendenti di incroci reciproci, proprio come aveva fatto Mendel. Malgrado questo, nell'espressione di quei caratteri c'erano ancora differenze sufficienti da poter distinguere quale dei due genitori li avesse trasmessi.

Analogamente, ricerche effettuate nel mio laboratorio e in quello di Alan C. Peterson del Ludwig Institute hanno dimostrato che parecchi caratteri genetici dei topi sono a mosaico. Se ci si domanda semplicemente se questi caratteri siano presenti, la risposta è affermativa, ma se ci si chiede con quale intensità essi vengano espressi, allora si notano gli effetti dell'imprinting parentale. Per alcuni caratteri, il grado di espressione può non essere molto importante, ma per altri può avere un'importanza fondamentale: per esempio, può voler dire essere affetto da tumore di Wilms o no, oppure manifestare la corea di Huntington all'età di cinque anni piuttosto che a 70.

Una delle aree di ricerca potenzialmente più fruttuose, nello studio dell'imprinting genomico, può essere l'ereditarietà di alcune malattie dell'uomo. Rimane ancora da vedere quanto questa impostazione della ricerca contribuirà alla comprensione delle malattie; i successi teorici, se non pratici, che abbiamo realizzato finora nella conoscenza della corea di Huntington e di alcuni tumori maligni che colpiscono l'infanzia ci danno però la speranza che un certo progresso potrà realmente essere acquisito.

BIBLIOGRAFIA

CATTANACH B. M. e KIRK M., *Differential Activity of Maternally and Paternally Derived Chromosome Regions in Mice* in «Nature», 315, n. 6019, 6 giugno 1985.

SOLTER D., *Differential Imprinting and Expression of Maternal and Paternal Genomes* in «Annual Reviews of Genetics», 22, pp. 127-146, 1988.

SCRABLE H. e altri, *A Model for Embryonal Rhabdomyosarcoma Tumorigenesis that Involves Genome Imprinting* in «Proceedings of the National Academy of Sciences», 86, n. 19, ottobre 1989.

SHARP PC E500. IL VOSTRO PENSIERO DISPONE DI 1101 MOSSE.

Chi opera in un ambito scientifico lavora molto con la testa e deve muovere continuamente il pensiero per aggirare ostacoli e saltare a nuove soluzioni. Ecco perchè Sharp vi suggerisce una semplice strategia per ampliare le mosse che già avete in mente fino a 1.101! Sono infatti 1.101 le funzioni tra matematica, ingegneria e statistica del PC E500, il cavallo vincente della gamma Pocket Computer, destinato a dare scaccomatto a qualsiasi vostro problema. Programmabile in BASIC, 124 costanti generali, 744 formule e dati, 233 funzioni di calcolo scientifico: un software completo per un pratico ed immediato utilizzo.

Grande display di 40 colonne per 4 linee, completamente grafico (240 x 32 punti), 32 Kb di memoria espandibili fino a 96 Kb. Funzione AER per registrare 99 equazioni personali. Possibilità di allacciamento a PC e periferiche, eccezionale velocità di elaborazione. Sharp PC E500. Il pezzo più professionale per muovere la vostra mente con rapidità in ogni direzione.

SHARP POCKET COMPUTER

melchioni

ADART C&C&P

La prospezione geochimica nell'esplorazione mineraria

La mobilizzazione nella crosta terrestre di alcuni elementi chimici forma aureole di dispersione facilmente rilevabili che consentono di localizzare minerali di notevole interesse, compresi i depositi di metalli preziosi

di Benedetto De Vivo

Dall'immediato dopoguerra a oggi si sono avuti notevoli sviluppi nel settore dell'esplorazione mineraria. Le compagnie hanno messo a punto nuove tecniche di prospezione, oltre a quelle specificamente geologiche che avevano portato alla scoperta dei giacimenti «a vista». (Essendo ormai stati individuati nel mondo tutti i giacimenti più evidenti, la prospezione «a vista» si trova a essere ormai del tutto superata: in Italia, per esempio, la maggior parte dei giacimenti attualmente conosciuti e sfruttati era già nota agli etruschi e ai romani.) Si sono così affinate le tecniche di prospezione geofisiche e geochimiche, fino a raggiungere un'importanza primaria ed entrare nella routine di qualunque programma di esplorazione mineraria. Secondo le stime dell'Association of Exploration Geochemists, per quanto riguarda la prospezione geochimica vengono raccolti annualmente otto milioni di campioni nei paesi non appartenenti all'ex blocco orientale e circa 10 milioni di campioni nella sola Unione Sovietica.

La prospezione geochimica si presta in particolare alla ricerca di giacimenti a basso tenore, specialmente di quelli situati in profondità e pertanto difficili da localizzare. Un esempio classico è dato dai cosiddetti *porphyry coppers*, depositi caratterizzati da basse concentrazioni di metalli, in particolare rame e molibdeno. Questo tipo di prospezione si basa su un concetto abbastanza intuitivo, probabilmente applicato istintivamente già dal primo «prospettore» in cerca di minerale utile. Questo il concetto: i giacimenti costituiscono una concentrazione anomala nelle rocce della crosta di uno o più metalli; nel corso di processi di alterazione delle rocce questi metalli entrano in parte nei prodotti dell'alterazione stessa e, meccanicamente o chimicamente, diffondono dal giacimento originando aureole di dispersione che possono essere scoperte con relativa facilità, in quanto hanno un'estensione di molto superiore a quella del giacimento.

L'analisi degli elementi minori e in traccia (gli elementi minori sono quelli con concentrazione compresa fra lo 0,1 e l'1 per cento; quelli in traccia sono inferiori allo 0,1 per cento) cominciò a costituire un ausilio alla ricerca «a vista» alla metà degli anni trenta. La nascita della prospezione geochimica come disciplina scientifica si deve al mineralista e petrografo di origine svizzera Victor Moritz Goldschmidt, il quale pubblicò nel 1933 un lavoro in cui si metteva in risalto come il carbone risulti spesso arricchito in numerosi elementi in traccia. Questo arricchimento sarebbe dovuto alla capacità posseduta dalle radici delle piante di estrarre elementi dal suolo, ovvero dal substrato sul quale vegetano. La successiva decomposizione delle piante rilascia nel suolo gli elementi che esse avevano estratto, cosicché i vari orizzonti del terreno risultano arricchiti in questi elementi.

Il lavoro di Goldschmidt aveva posto le basi della prospezione geochimica, una tecnica che, dopo il 1933, si sviluppò notevolmente in Unione Sovietica e, dal dopoguerra in poi, nei paesi occidentali e, in particolare, negli Stati Uniti e in Canada.

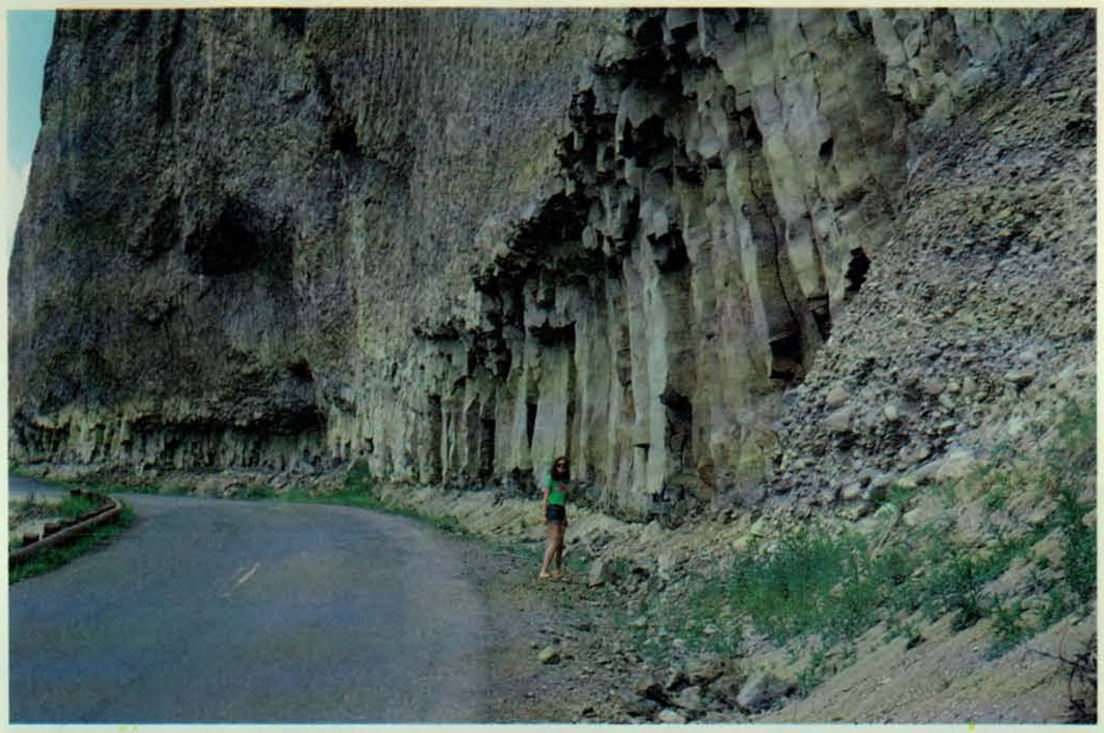
I metodi di prospezione geochimica si basano sull'analisi sistematica su larga scala della dispersione degli elementi attorno ai giacimenti minerari, dispersione che avviene a opera di agenti chimico-fisici. Il processo di dispersione può riguardare due tipi di ambiente geochimico, uno primario e uno secondario.

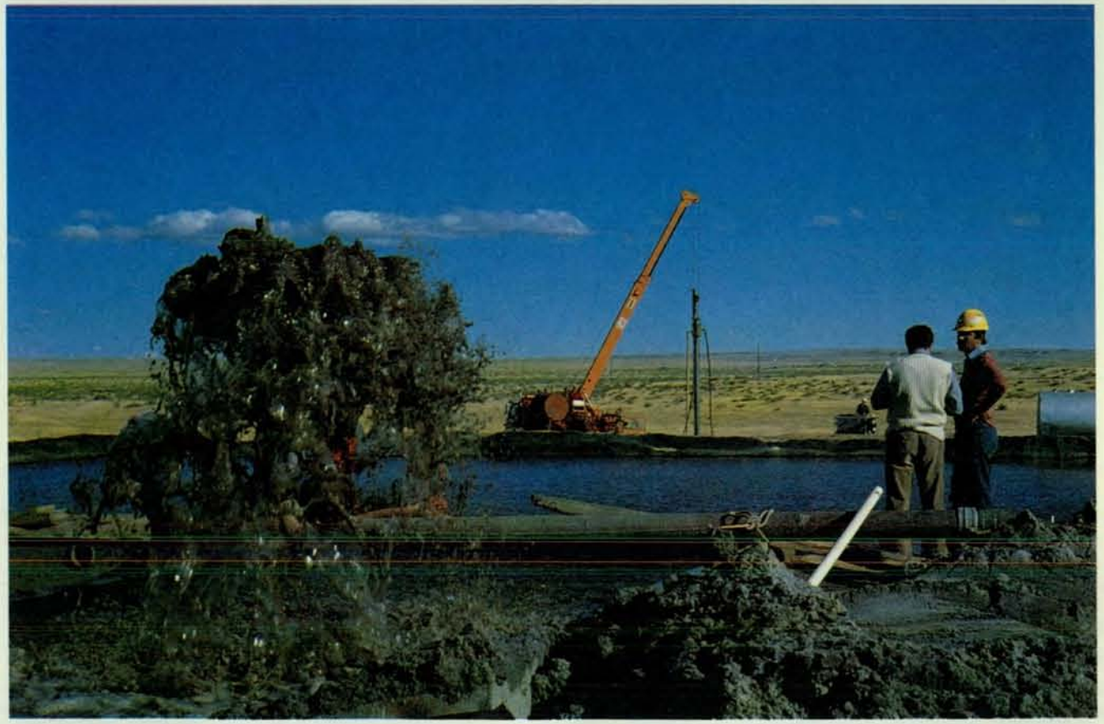
L'ambiente primario è caratterizzato normalmente da condizioni di temperatura e pressione elevate e comprende la zona che si estende dal livello inferiore delle acque meteoriche circolanti a quello più profondo in cui avvengono i processi di metamorfismo e di differenziazione magmatica. In questo ambiente vi è normalmente poco ossigeno libero e la circolazione dei fluidi è limitata.

L'ambiente secondario è invece caratterizzato da condizioni di bassa temperatura e pressione; in esso si hanno processi di alterazione superficiale, di formazione dei suoli e di sedimentazione. In questo ambiente vi è abbondanza di ossigeno molecolare e di altri gas - particolarmente anidride carbonica - e libera circolazione di fluidi.

Il movimento di materiale fra ambiente primario e ambiente secondario può essere considerato alla stregua di un processo ciclico chiuso che va dalle rocce ignee alle rocce sedimentarie: è questo il cosiddetto «ciclo geochimico» (*si veda l'illustrazione a pagina 30*). La risposta di un elemento ai processi di dispersione dipende dalla mobilità di quell'elemento, vale a dire dalla facilità con cui esso è in grado di spostarsi all'interno di un ambiente specifico.

In ambiente primario, durante la cristallizzazione dei magmi, i costituenti che non possono in tutto o in parte entrare nella struttura cristallina dei minerali (in generale silicati) si concentrano nel fluido residuo mobile, che è costituito prevalentemente da acqua e si trova in forma dapprima gassosa e successivamente liquida. Con il progressivo diminuire della temperatura, questi fluidi residui tendono ad assumere proprietà di soluzioni acquose. Da queste soluzioni, dette idrotermali, possono a questo punto del ciclo geochimico primario formarsi mineralizzazioni per precipitazione. Allo stesso modo durante i processi di metamorfismo possono avere origine soluzioni idrotermali. Si possono ricavare informazioni sulla mobilità degli ele-





Il bacino del Wyoming rappresenta una delle aree più ricche della Terra dal punto di vista giacimentologico, soprattutto riguardo alla mineralizzazione di uranio, grazie a un'intensa attività vulcanica e idrotermale. Nella fotografia in alto una colata ignimbritica nel Parco nazionale di Yellowstone; in basso la miniera di Lake dove l'uranio è «coltivato» con la tecnica di lisciviazione per soluzione.

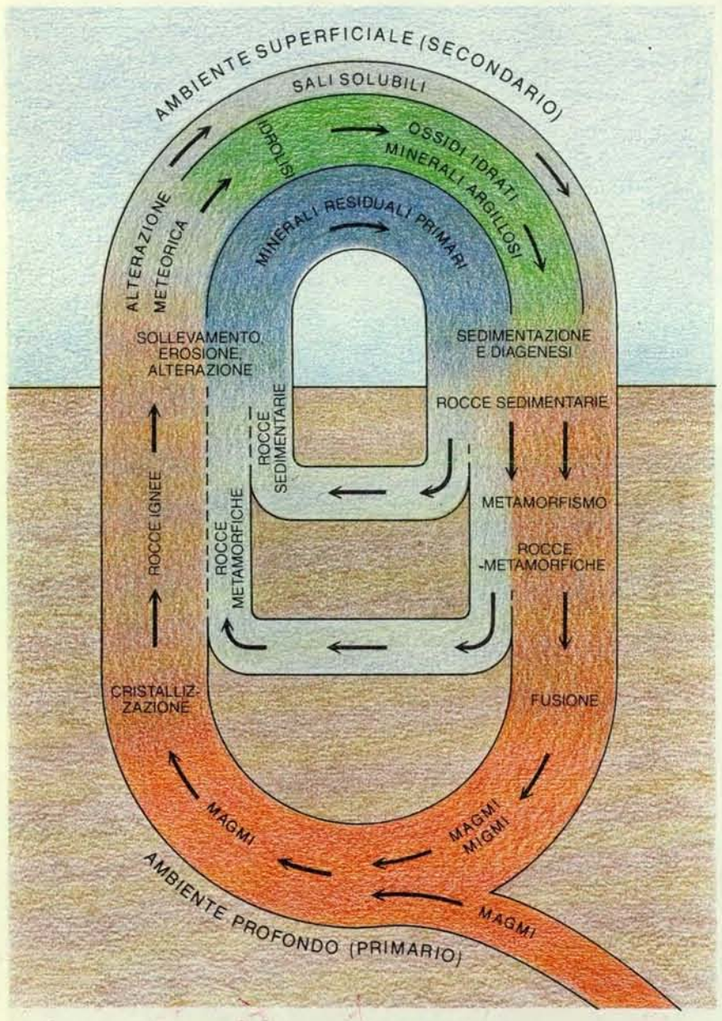
menti osservando il progressivo impoverimento o arricchimento nelle specie minerali che si formano nei diversi stadi di un ciclo di differenziazione.

Un esempio di dispersione geochimica e quindi di mobilità in ambiente primario è dato dal processo di alterazione delle rocce incassanti, noto in letteratura come «alterazione idrotermale». Il suo studio trova ormai larga applicazione nella ricerca di giacimenti minerari, visto che l'alterazione idrotermale rappresenta una prova dell'esistenza di soluzioni idrotermali. Nella ricerca dei *porphyry coppers* viene largamente utilizzato il modello di alterazione proposto nel 1970 da James D. Lowell e John M. Guilbert dell'Università dell'Arizona. Questi ricercatori hanno riconosciuto l'esistenza di strette relazioni tra i giacimenti tipo *porphyry coppers* e la zonazione laterale e verticale di associazioni mineralogiche derivanti da alterazione idrotermale di tipo potassico, fillico, argillitico e propilitico (*si veda l'illustrazione a pagina 31*).

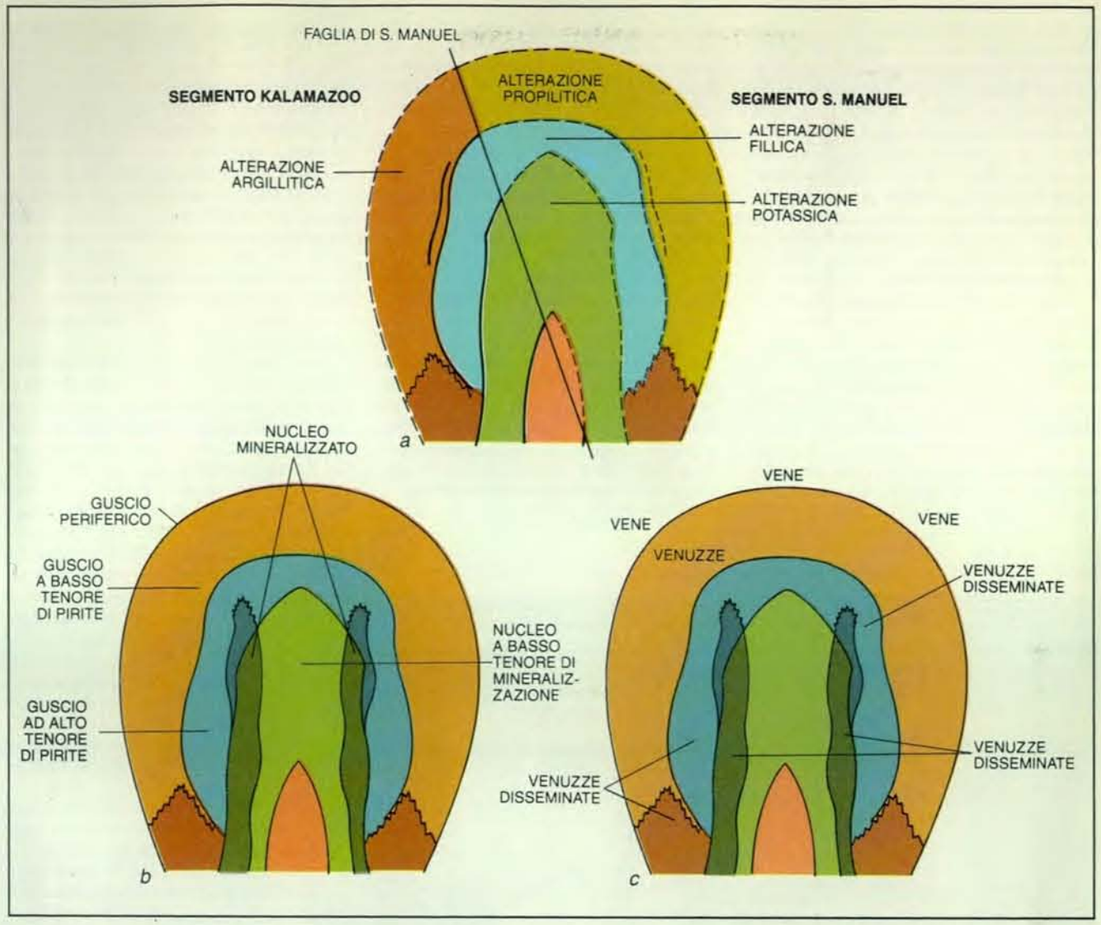
La formazione delle aureole di dispersione secondarie è dovuta alla mobilità degli elementi nell'ambiente secondario. Il processo è diverso da quello che si riscontra nell'ambiente primario, e diversi sono i meccanismi di formazione delle aureole. Infatti il fenomeno viene in questo caso attivato dall'alterazione chimico-fisica delle rocce.

Tra i diversi fattori che influenzano la mobilità degli elementi nell'ambiente secondario hanno un ruolo fondamentale le variazioni di *E*h e di *p*H. Il *p*H esprime notoriamente l'acidità o l'alcalinità di una soluzione acquosa e sta a indicare la concentrazione di ioni H^+ e OH^-. È definito come il cologaritmo decimale della concentrazione degli ioni idrogeno. L'*E*h di una soluzione dà invece la misura della tendenza ossidante o riducente di un sistema e, poiché l'ossidoriduzione implica un trasferimento di elettroni da un elemento a un altro, *E*h viene espresso in volt o millivolt. *E*h e *p*H determinano quindi la concentrazione di uno ione in soluzione e il mutamento di uno di questi due parametri basta a causarne la precipitazione o l'ulteriore soluzione. Per la stretta interdipendenza fra i due parametri, i diagrammi *E*h-*p*H trovano largo impiego nelle diverse branche della geochimica (*si veda l'illustrazione a pagina 32*).

Un altro fattore che influenza molto le condizioni di mobilità di un elemento in ambiente secondario è l'adsorbimento. Per quanto riguarda in particolare appunto l'adsorbimento, gli ossidi idrati di ferro e di manganese hanno un ruolo di primaria importanza nella prospezione geochimica. Durante la precipitazione (in condizioni ossidanti) essi possono infatti incorporare per coprecipitazione altri elementi che, normalmente, non sono interessati dalle variazioni di *E*h e di *p*H e hanno inoltre la tendenza ad adsorbire gli elementi con cui vengono a contatto. Ciò nuoce alla prospezione geochimica, sia perché viene ridotta la mobilità degli elementi nell'ambiente secondario (e le aureole di dispersione risultano quindi limitate), sia perché si formano false anomalie geochimiche. I fenomeni di adsorbimento sono altresì determinati dalla presenza di materia organica e dall'abbondanza di minerali argillosi. La mobilità degli elementi in ambiente secondario, oltre a dipendere da variazioni dei parametri *E*h e *p*H e dai fenomeni di adsorbimento, è funzione anche di altri fattori, tra i quali vanno ricordati: natura del mezzo di trasporto, meccanismo di trasporto, variabilità dei tipi litologici, presenza di microrganismi, formazione di ioni complessi, presenza di gas disciolti, permeabilità e porosità delle rocce, viscosità e velocità dei fluidi.



Il trasferimento di materia da un ambiente crostale all'altro può essere considerato un processo ciclico come quello illustrato da questo schema elaborato nel 1979 da A. W. Rose, H. E. Hawkes e J. S. Webb. A partire dalle rocce ignee primarie si arriva, per erosione e alterazione superficiale, alla formazione di rocce sedimentarie. Da queste, se interviene un aumento di temperatura e pressione, derivano per metamorfismo altri tipi di rocce e fluidi idrotermali. Quando, per l'erosione, le rocce si trovano di nuovo esposte agli agenti atmosferici, si ha una ridistribuzione nell'ambiente superficiale degli elementi che le compongono con formazione di nuove rocce sedimentarie. In questo modo il ciclo geochimico si chiude. Alcune tappe del ciclo possono venire saltate: per esempio le rocce sedimentarie possono subire alterazione meteorica senza passare attraverso il processo metamorfico.



Schema delle zone di alterazione e mineralizzazione nel giacimento di San Manuel-Kalamazoo in Arizona, secondo J. D. Lowell e J. N. Guilbert. In *a* si vedono i tipi di alterazione; le linee tratteggiate indicano incertezza sulla continuità e la localizzazione delle zone. In *b* appaiono le zone mineralizzate, mentre in *c* è mostrata la distribuzione dei solfuri. L'alterazione propilitica si estende 700-800 metri oltre la zona mineralizzata. Lungo lo stesso intervallo di alterazione idrotermale i solfuri vanno, dall'interno verso l'esterno, da associazioni tipo calcopirite-molibdenite-pirite ad associazioni tipo blenda-galena con quantità subordinate di oro e argento.

Di estrema utilità nella prospezione geochimica è il fatto che alcuni elementi tendono a ritrovarsi associati nelle rocce della crosta, tendenza interpretabile semplicemente come la manifestazione di una mobilità relativa simile da parte di elementi diversi, in risposta alle variazioni ambientali. Ne consegue che la presenza di un elemento appartenente a una certa associazione suggerisce l'esistenza di altri elementi associati. Per questo si ricorre, nella prospezione, agli indicatori geochimici: elementi relativamente mobili che si ritrovano in stretta associazione con altri elementi. Questi indicatori sono rilevabili più facilmente, o perché formano aureole di dispersione più ampie, oppure perché è più agevole registrarli con tecniche strumentali. Così, per esempio, il mercurio e l'arsenico sono indicatori della presenza di giacimenti auriferi; il selenio, il vanadio e il molibdeno indicano invece giacimenti di uranio epigenetico (*si veda la tabella a pagina 34*).

Per concludere, la prospezione geochimica nell'esplorazione mineraria ha lo scopo di evidenziare zone anomale, nelle quali cioè si rinvengono concentrazioni particolarmente abbondanti di elementi chimici riconducibili alla presenza di un corpo mineralizzato. Essa è diretta a determinare la distribuzione degli elementi nelle acque superficiali e sotterranee (prospezione idrogeochimica), nei vegetali (prospezione biogeochimica), nei suoli (prospezione pedogeochimica), nelle rocce (prospezione litogeochimica), nei sedimenti fluviali e nelle esalazioni gassose.

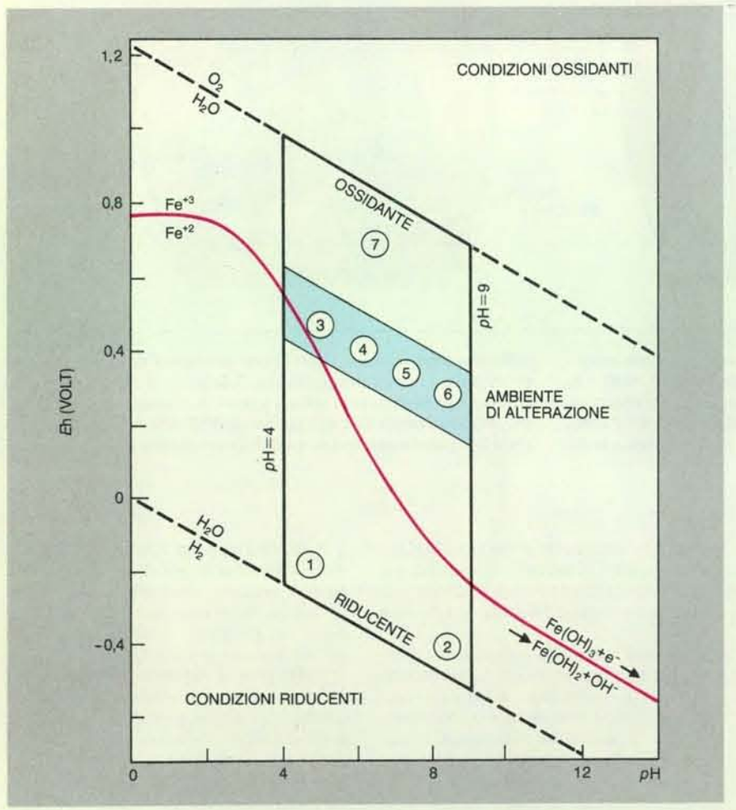
Definiti i principi fondamentali della prospezione geochimica in senso stretto, vediamo ora come è organizzata la ricerca mineraria, qual è il ruolo specifico del geologo e di quali strumenti si avvale durante le varie fasi della ricerca.

Nella fase di organizzazione, occorre formare un'équipe con competenze specifiche in geologia, geofisica, geochimica e nei rispettivi metodi di prospezione; si devono poi scegliere gli obiettivi di potenziale interesse minerario, formulare un programma di esplorazione con stima dei costi (comprensivo delle fasi preliminari di studio, di una prima fase riconoscitiva e delle varie fasi di prospezione) e dirigere infine le varie fasi della ricerca, interpretando i dati in modo da orientarne lo sviluppo in relazione alla convenienza economica.

Una volta individuato l'obiettivo su cui dirigere la ricerca e dopo avere eseguito gli studi preliminari, ha inizio la ricerca in senso stretto, nella quale si distinguono una fase riconoscitiva, una di esplorazione e quella finale di valorizzazione. Nella prima fase lo scopo è quello di circoscrivere le aree di maggiore interesse minerario, facendo prospezioni su larga scala per mezzo di telerilevamento, fotogeologia, rilevamento geologico riconoscitivo, prospezione aeromagnetica e prospezione geochimica regionale.

Nella fase di esplorazione si tratta di delimitare le aree scelte nella fase riconoscitiva determinandone i caratteri generali. In questa seconda fase gli strumenti sono il rilevamento geologico, la prospezione geofisica e geochimica di dettaglio, l'analisi petrografica, stratigrafica e strutturale e, inoltre, la distribuzione dei minerali di alterazione e dei corpi rocciosi potenzialmente mineralizzati. Nella fase di valorizzazione si eseguono perforazioni sistematiche per definire la grandezza, la configurazione e, approssimativamente, il tenore del giacimento. Alla ricerca fanno seguito le operazioni di sfruttamento.

In fase esplorativa è importante per il geologo avere una conoscenza dei fenomeni che presiedono alla formazione dei giacimenti minerari (nel contesto della dinamica crostale) e inquadrare i fenomeni metallogenetici alla luce degli sviluppi della teoria della tettonica a zolle. Data la mancanza di indizi superficiali, l'esplorazione viene in un primo momento orientata verso la ricerca di ambienti favorevoli all'adunamento di giacimenti (ambiente vulcanogenico, di apertura oceanica, di subduzione, di mobilizzazione cratonica, sedimentario-evaporitico, singenetico lagunare-marino, di acque connato-meteoriche, di concentrazioni residuali e metamorfico).

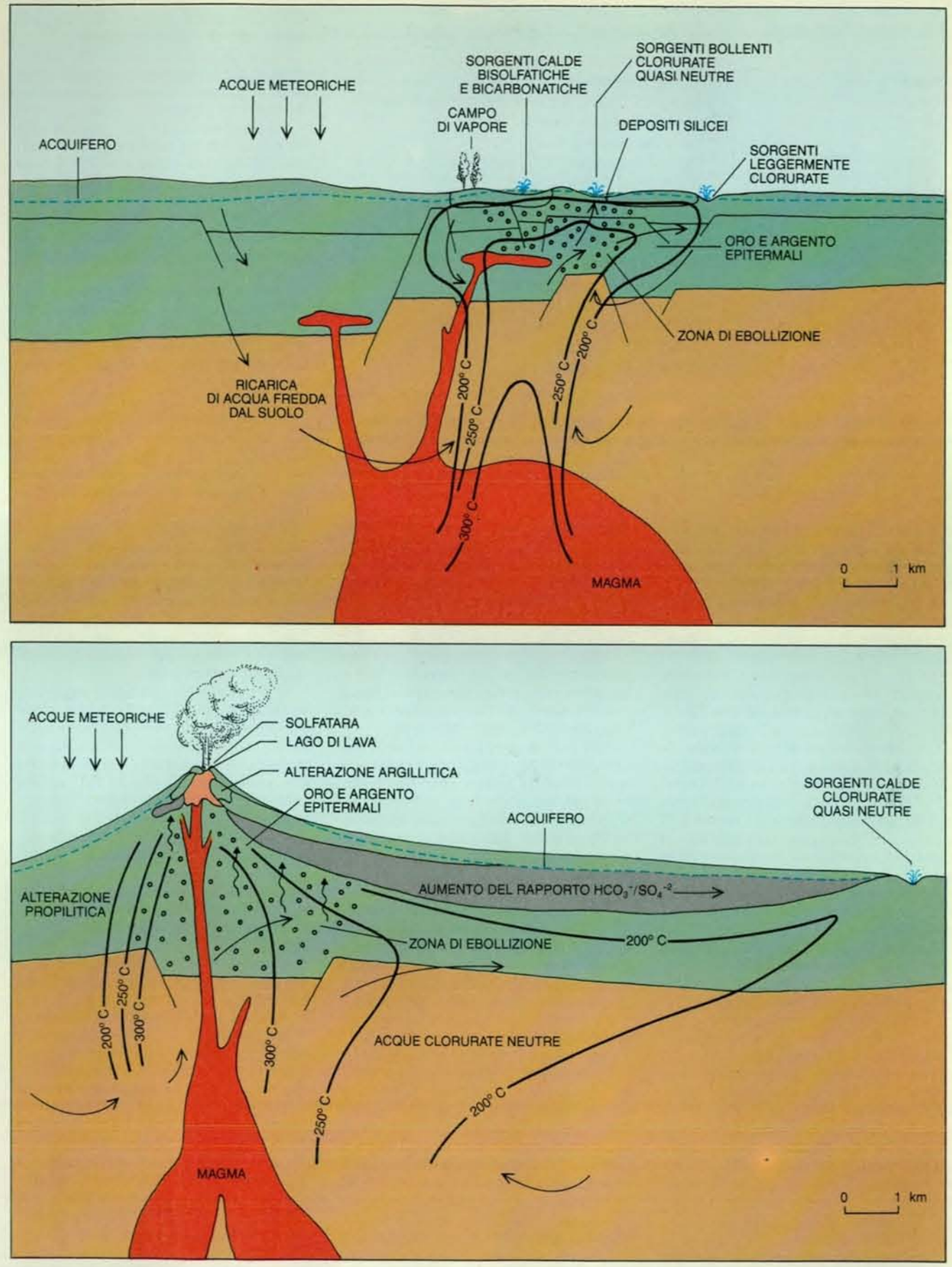


In questo diagramma sono indicati i limiti dei due parametri E*h*-*p*H negli ambienti superficiali. Il limite superiore dell'ambiente di alterazione è a contatto diretto con l'atmosfera, mentre il limite inferiore corrisponde a quello della falda acquifera. La posizione dei numeri nel cerchietto corrisponde alle condizioni di E*h* e *p*H di altri ambienti: 1) paludi e suoli; 2) sedimenti marini riducenti; 3) acque acide di miniera; 4) pioggia; 5) acque fluviali; 6) acque oceaniche; 7) depositi a solfuri di piombo ossidanti. Il potenziale di ossidoriduzione per ioni e idrossidi di ferro è dato per la temperatura di 25 °C e per una mole.

L'esplorazione mineraria non va intesa soltanto come finalizzata alla scoperta di minerali economicamente sfruttabili in tempi brevi (cioè del giacimento propriamente detto), ma anche come strumento per effettuare una stima delle risorse disponibili in un certo territorio. La differenza fondamentale tra risorse e riserve consiste nel fatto che le riserve sono economicamente sfruttabili fin dalla scoperta, mentre le risorse rappresentano solo un potenziale economico. Le risorse possono diventare riserve al variare delle condizioni economiche che determinano il valore di mercato di un certo minerale.

Ogni paese moderno ha interesse a stimare attraverso i propri organi di ricerca (generalmente i servizi geologici di stato) le risorse disponibili sul proprio territorio. Una valutazione delle risorse deve necessariamente passare attraverso tre stadi. Nel primo si procede all'identificazione dei tipi di depositi che possono essere presenti all'interno di un certo territorio, sulla base delle caratteristiche geologiche del territorio stesso. Nel secondo stadio si delimitano, all'interno di quel territorio, le aree che possono contenere particolari tipi di giacimenti. Ciò è possibile sulla base di un esame critico di tutti i dati geologici disponibili e interpretabili alla luce di modelli metallogenetici noti. Il terzo stadio consiste nello stimare quanti giacimenti non ancora scoperti potrebbero esistere all'interno del territorio considerato.

L'informazione chiave che il geologo deve possedere per procedere con maggiore probabilità di successo attraverso i tre stadi sopracitati è quella concernente le caratteristiche dei diversi tipi di deposito minerario. Questa informazione deve essere organizzata in modo non ambiguo e dare il più possibile risposte «oggettive». In questo senso le caratteristiche dei vari modelli di deposito minerario riportati dai due ricercatori statunitensi Dennis P. Cox e Donald P. Singer dello US Geological Survey assolvono egregiamente al compito. I modelli riportati da Cox e Singer nel loro lavoro sono il frutto di dati derivanti dalla catalogazione di 3900 giacimenti minerari e sono rappresentativi di quasi tutti i tipi di giacimento in cui ci si può imbattere. Per ogni modello è riportata una breve descrizione del contesto geologico in cui il giacimento si trova (tipi di rocce, età, struttura, tettonica, ambiente di deposizione, tipi di depositi associati); viene descritta la mineralogia, la tessitura delle mineralizzazioni, i tipi di rocce alterate, le caratteristiche dell'alterazione idrotermale, nonché le caratteristiche geochimiche per ogni modello di deposito minerario. Vengono inoltre fornite, sotto forma di tabelle e diagrammi, informazioni sul tonnellaggio e sul tenore, sui tipi di deposito che potrebbero trovarsi associati in particolari ambienti geologici, sugli elementi primari e accessori, sulle anomalie geochimiche riscon-



Le illustrazioni (tratte da R. W. Henley e A. J. Ellis) schematizzano la configurazione di sistemi geotermici, con l'indicazione degli ambienti di deposizione delle mineralizzazioni a oro e argento epitermali. In basso si vede in particolare come la topografia accentuata dei vulcani andesitici provochi uno spostamento laterale di soluzioni calde ricche in cloruri fino a 20 chilometri dal centro vulcanico.

ELEMENTI INDICATORI	TIPO DI DEPOSITO
As	Au, Ag; tipo filoniano
As	Au-Ag-Cu-Co-Zn; giacimenti a solfuri complessi
B	W-Be-Zn-Mo-Cu-Pb; skarn
B	Sn-W-Be; filoni o greisen
Hg	Pb-Zn-Ag; depositi a solfuri complessi
Mo	W-Sn; depositi metamorfici di contatto
Mn	Ba-Ag; depositi filoniani; porphyry coppers
Se, V, Mo	U; tipo arenarie
Cu, Bi, As, Co, Mo, Ni	U; tipo filoniano
Mo, Te, Au	Porphyry coppers
Pd, Cr, Cu, Ni, Co	Platino in rocce metamorfiche
Zn	Ag-Pb-Zn; depositi a solfuri in generale
Zn, Cu	Cu-Pb-Zn; depositi a solfuri in generale
Rn	U; depositi di ogni tipo
SO_4^{-2}	Depositi a solfuri di tutti i tipi

Svariati elementi chimici di grande mobilità (in grado quindi di originare aureole di dispersione ampie e facilmente rilevabili) fungono da indicatori per alcuni tipi di deposito.

trabili per i diversi tipi di giacimento e infine sui minerali che costituiscono il giacimento, la ganga e la roccia incassante (con le relative alterazioni).

Il procedimento che conduce a delimitare una porzione del territorio come più probabile per la scoperta di nuovi giacimenti può a sua volta essere suddiviso in tre fasi: in una prima fase si identifica il tipo di deposito potenzialmente presente; successivamente si effettua l'identificazione preliminare delle unità geologiche cartografate e di tutte le altre caratteristiche geologiche che determinano la mineralizzazione del tipo di deposito ricercato; infine, si utilizzano tutte le informazioni storiche e i risultati di studi geochimici, geologici e geofisici di dettaglio, allo scopo di eliminare tutte le aree che sicuramente non sono adatte a contenere quel tipo di deposito minerario, per delimitare viceversa quelle con potenziali risorse.

La ricerca, nella prima fase, procede per analogia, per estrapolazione o per associazione. Per analogia vengono presi in considerazione i tipi di deposito reperibili in aree con caratteristiche geologiche simili a quelle dell'area appunto investigata. L'estrapolazione viene soprattutto utilizzata in aree non bene esplorate, adiacenti ad altre ben conosciute con caratteristiche geologiche simili. L'associazione si utilizza fra tipi di deposito strettamente apparentati; per esempio, dato che i depositi di barite sono associati a depositi di piombo e zinco, la presenza di uno dei due tipi di deposito suggerisce la possibilità che anche l'altro tipo sia presente.

Nell'esplorazione mineraria, arrivare a definire con certezza le aree sterili per specifici tipi di deposito minerario è comunque un risultato importantissimo, perché significa risparmiare risorse e capitali in ricerche future. Anche in questo caso, comunque, il termine certezza è alquanto relativo, nel senso che è, per così dire, funzione di modelli metallogenetici validi attualmente. L'affacciarsi di nuovi modelli sulla scena potrebbe sconvolgere tutte le filosofie di ricerca. Un esempio in questo senso può essere rappresentato dall'enorme impulso che negli ultimi anni, su scala mondiale, ha assunto la ricerca dell'oro associato a sistemi epitermali in rocce vulcaniche.

Alla luce di questi nuovi modelli metallogenetici per metalli preziosi, tutte le aree vulcaniche italiane potrebbero avere un potenziale aurifero. Nel paragrafo seguente verrà illustrato in che modo potrebbero essere delimitate le aree vulcaniche atte a contenere risorse aurifere non ancora scoperte.

Vediamo in primo luogo quali sono le caratteristiche generali dei depositi epitermali in rocce vulcaniche subaeree; cercheremo poi di definire quali aree, in Italia, potrebbero essere atte a contenere mineralizzazioni di tipo epitermale.

Le rocce vulcaniche si ritrovano comunemente lungo i margini di zolla. I campi vulcanici ospitano una grande varietà di depositi minerari, raggruppati come depositi epitermali. Si tratta di mineralizzazioni, appunto, di ambiente vulcanico superficiale, formatesi a temperature comprese fra i 50 e i 200 gradi Celsius. Tra queste ricadono principalmente le mineralizzazioni a mercurio, ad antimonio, a oro sotto forma di tellururi con alunite associata e sotto forma di seleniuri, ma anche a rame, piombo, zinco, stagno e uranio. La profondità massima di formazione di questo tipo di depositi non è superiore ai 1000 metri, ragion per cui nei terreni vulcanici che sono stati erosi al di sotto di questa profondità risulta inutile mettersi alla ricerca di questo tipo di mineralizzazioni.

Nel modello di deposizione dei metalli preziosi epitermali, i sistemi geotermici attivi in terreni vulcanici sono considerati come un laboratorio naturale in cui si possono riscontrare tutti i processi che hanno dato luogo in passato alla formazione di depositi epitermali. Per questo motivo lo studio dei fluidi idrotermali associati a un sistema geotermico attivo riveste notevole importanza non solo ai fini delle ricerche geotermiche per la produzione di energia, ma anche per le ricadute che se ne hanno nel campo dell'esplorazione mineraria. In questo senso il contributo che viene dallo studio delle inclusioni fluide nei minerali idrotermali dei campi geotermici riveste un ruolo di primaria importanza. In altre parole, i depositi epitermali di metalli preziosi ospitati in rocce «antiche» non sono altro che il frutto di sistemi geotermici fossili (*si vedano le illustrazioni a pagina 33*).

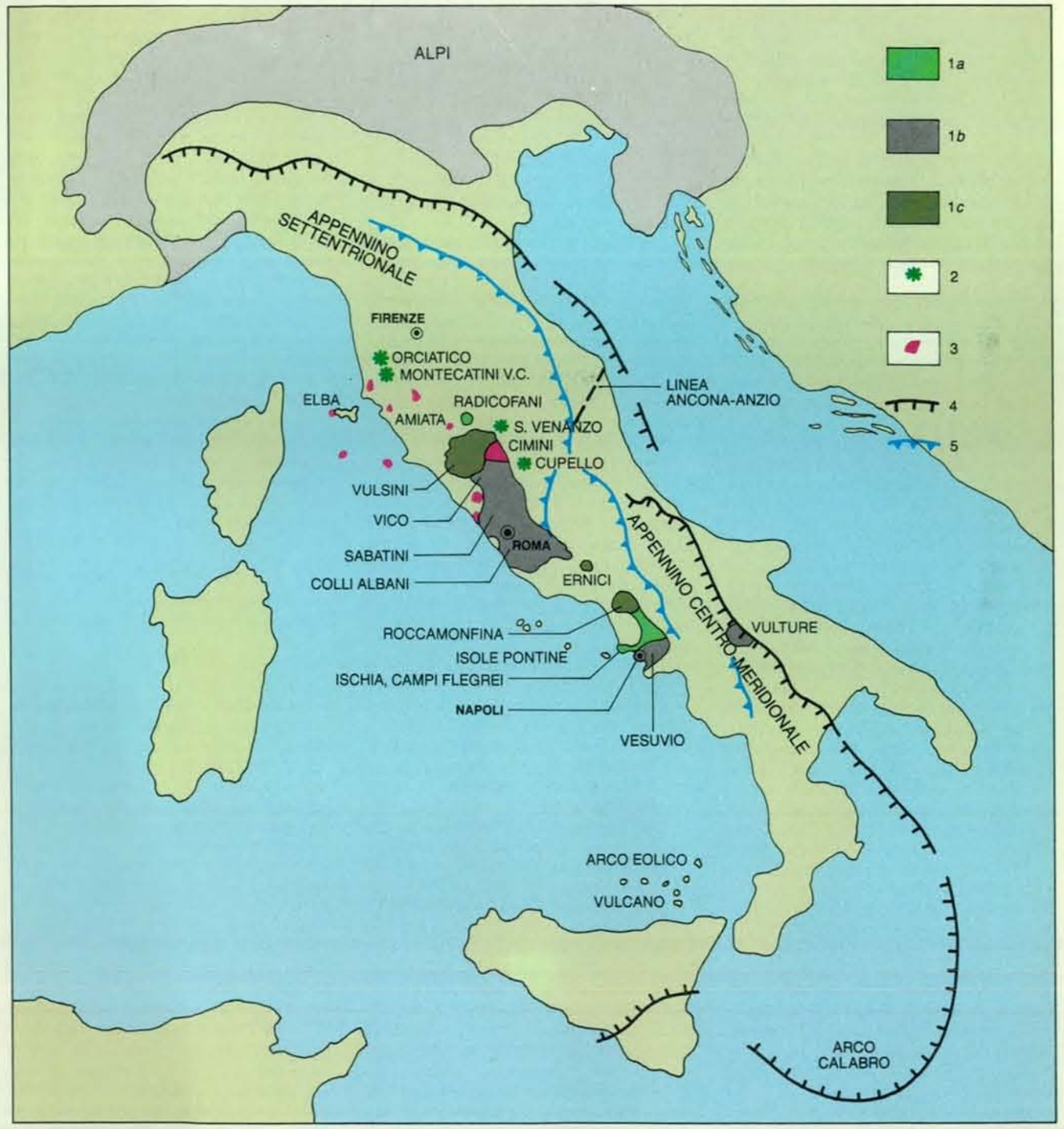
I processi geotermici più importanti che influenzano la deposizione di mineralizzazioni sono la circolazione di acque calde clorurate alcaline, l'ebollizione e il raffreddamento di queste acque in ambiente prossimo alla superficie, il mescolamento delle acque clorurate alcaline ascendenti con acque meteoriche ossidanti, da neutre ad acide, discendenti. Tutti questi processi determinano altresì una caratteristica mineralogia di alterazione delle rocce ospiti.

È probabile che l'origine degli ioni metallici trasportati nelle acque calde ascendenti risieda nelle rocce del basamento che vengono lisciviate dalle acque meteoriche discendenti, a mano a mano che queste si riscaldano ed entrano nel sistema convettivo delle acque clorurate alcaline. Queste acque possono comunque ricevere un contributo anche da parte di fluidi di origine magmatica, in considerazione del fatto che la solubilità di alcuni metalli è controllata in modo rilevante dal chimismo e dalla temperatura di queste acque calde ascendenti.

Le caratteristiche geologiche utili a definire un programma di esplorazione mineraria e che controllano l'esistenza di depositi di metalli preziosi epitermali comprendono il tipo di rocce, l'età dei terreni ospitanti le mineralizzazioni, il tipo di vulcanismo, l'ambiente tettonico e il tipo di mineralizzazione associato. Come già detto, i depositi epitermali si trovano in rocce vulcaniche subaeree (basalti, andesiti, daciti, rioliti e sedimenti vulcanoclastici) di età terziaria o più recente, prevalentemente ad affinità da calc-alcalina ad alcalina e con vulcanismo bimodale riolitico-basaltico. È fondamentale la presenza di un centro intrusivo (magma silicico) abbastanza superficiale, che faccia da motore per la messa in moto di un sistema idrotermale convettivo, capace di trasportare e depositare metalli. Perché questo si verifichi è necessaria la presenza di una porosità e permeabilità primaria in tufi non saldati, in sedimenti vulcanoclastici e in colate di breccia. L'ambiente tettonico dei depositi epitermali comprende archi insulari e continentali, faglie trasformi lungo margini continentali, nonché zone di *rifting* continentale (ossia zone dove la crosta si frattura). I tipi di giacimenti, infine, generalmente associati ai depositi epitermali di metalli preziosi sono depositi di mercurio, antimonio, fluorite e perlite.

Per quanto riguarda le caratteristiche dei depositi mineralizzati, assume un ruolo importante la presenza di porosità secondaria dovuta a faglie e fratture, mentre la mineralogia è espressa da galena, blenda, argentite, oro, tetraedrite, pirite, calcopirite, elettro, tellururi d'oro, bornite, arsenopirite, enargite, luzonite, covellite, calcocite, stibnite, realgar e orpimento. I minerali di ganga sono: quarzo, calcedonio, opale, calcite, barite, celestite, fluorite, clorite, rodocrosite, adularia, sericite, caolinite, alunite, siderite, ankerite, jarosite ed ematite.

L'alterazione idrotermale è la prova



È indicata la distribuzione (secondo A. Peccerillo e P. Manetti) del vulcanismo potassico in Italia, così ripartibile: 1) provincia romana suddivisa in 1*a*, serie potassica; 1*b*, serie alta in potassio; 1*c*, serie potassica e serie alta in potassio; 2) centri isolati iperpotassici; 3) rocce anatettiche crostali della provincia toscana; 4) limiti esterni della catena appenninica; 5) limiti esterni delle falde principali.

più diretta dell'esistenza di un sistema idrotermale attivo nell'area studiata. Le rocce ospiti possono presentare tracce di silicizzazione (quarzo, calcedonio, opale), argillificazione (caolinite, illite, sericite, montmorillonite), propilitizzazione (clorite, calcite, pirite), argillificazione avanzata (alunite, silice, zunyite, diasporo, pirofillite, caolinite, corindone), e ossidazione (ematite, jarosite, goethite, alunite).

Quello che più conta dal nostro punto di vista è il fatto che i depositi epitermali sono sempre contrassegnati da anomalie geochimiche rappresentate da associazioni di elementi quali oro-argento-antimonio-arsenico-mercurio-tallio; oro-argento-tellurio-fluoro-arsenico; oro-argento-arsenico-tungsteno; oro-argento-arsenico-manganese-zinco-piombo-rame. Queste associazioni sono zonate sia verticalmente sia orizzontalmente rispetto all'area mineralizzata per cui il loro riconoscimento è una guida importante per una corretta interpretazione del tipo di mineralizzazione e per una sua precisa localizzazione.

Definite le caratteristiche generali dei depositi di metalli preziosi epitermali, cerchiamo di rispondere alla domanda: esistono nel territorio italiano presupposti geologici e indizi di natura geochimica e metallogenica che possano incoraggiare ricerche per l'individuazione di risorse aurifere epitermali?

Esaminiamo prima il tipo di vulcanismo italiano. Il bordo tirrenico della penisola italiana è caratterizzato da vulcanismo, da recente ad attivo, alcalino potassico, che costituisce la ben nota provincia comagmatica romana. Due petrologi italiani, Angelo Peccerillo dell'Università di Messina e Piero Manetti dell'Università di Firenze, sulla base di dati petrologici e geochimici indicano la presenza di quattro gruppi di rocce potassiche: 1) quelle della cosiddetta serie potassica, costituita da rocce con composizione da trachibasalti a trachiti, quasi sature in silice, con un rapporto K_2O/Na_2O prossimo a 1; 2) quelle della serie «alta in potassio», costituita da rocce con composizione dalle tefriti a leucite fino alle fonoliti a leucite, fortemente sottosatura in silice, con alti rapporti K_2O/Na_2O e arricchita in elementi quali cesio, rubidio, bario, torio, uranio; 3) quelle dei magmi iperpotassici, costituite da rocce note come melititi a leucite e kalsilite, più fortemente sottosature in silice, con rapporto K_2O/Na_2O molto alto e infine (4) le rocce note come selagiti (*minettes*), leggermente sovrasature in silice. Le rocce iperpotassiche sono volumetricamente subordinate rispetto alle serie potassiche e alte in potassio (*si veda l'illustrazione a pagina 35*).

In base a quanto si è detto prima, se il livello di erosione non supera i 1000 metri, è possibile la presenza di depositi epitermali auriferi (depositi a oro-argento e depositi auriferi a quarzo-alunite). Inoltre, in funzione del tipo di basamento, si possono ritrovare depositi epitermali: «tipo Creede» se il basamento è costituito da rocce sedimentarie comprendenti sequenze di calcari ed evaporiti, oppure sedimenti che hanno intrappolato acque marine; «tipo Comstock» se il basamento è costituito da rocce sedimentarie clastiche o da loro equivalenti metamorfiche; «tipo Sado» se il basamento è costituito da rocce vulcaniche più antiche oppure da rocce intrusive.

Gli elementi da prendere in considerazione per avere informazioni sulla potenzialità del territorio italiano in quanto a risorse di depositi epitermali sono anzitutto la presenza di terreni vulcanici favorevoli e poi il tipo di basamento, il grado di erosione, la presenza di centri intrusivi, di sistemi idrotermali, di mineralizzazioni o di impronte geochimiche che fungano da traccianti e infine l'esistenza di porosità primaria o secondaria.

I nostri terreni vulcanici sono ampiamente favorevoli e rispondono appieno alle caratteristiche generali che abbiamo prima enunciato. Nell'ambito dei terreni vulcanici sono da privilegiare quelli dominati da sequenze tufacee e lave andesitiche, dacitiche e riolitiche, rispetto a quelli dominati da basalti o andesiti basaltiche. Questa prima selezione può essere eseguita semplicemente utilizzando la cartografia geologica esistente.

Il tipo di basamento presente nelle varie aree vulcaniche italiane può essere utile a darci indicazioni sul tipo di deposito epitermale che può essere presente. In Italia ci si dovrebbero aspettare depositi «tipo Creede», considerato che le sequenze carbonatiche ed evaporitiche sono certamente quelle dominanti nel basamento interessato dalle serie vulcaniche.

Tutte le aree vulcaniche con un intenso grado di erosione (oltre i 1000 metri) andrebbero eliminate dalle aree di potenziale interesse. Per tutte o quasi le aree vulcaniche italiane questo grado di erosione non è stato certamente raggiunto, trattandosi di vulcanismo recente o ancora attivo.

L'esistenza di centri intrusivi è largamente testimoniata dalla presenza di strutture calderiche, duomi di lava ed evidenze geofisiche.

La presenza di sistemi idrotermali attivi è documentata ovviamente dai vari campi geotermici, come quelli di Larderello-Travale-Radicofani, in Toscana; quelli dell'alto Lazio, dei Campi Flegrei e di Vulcano nelle isole Eolie. L'esistenza di un sistema idrotermale fossile in aree vulcaniche non più attive può essere evidenziata mediante ricerche geochimico-geologiche in cui assumono un ruolo fondamentale gli studi sulle inclusioni fluide contenute in minerali di rocce alterate, attraverso un attento rilevamento del tipo di alterazione idrotermale, nonché attraverso l'uso di tecniche geofisiche. Aree con sistemi idrotermali fossili sono in particolare quelle della Sardegna occidentale, dove è presente un magmatismo calc-alcalino di età compresa fra 19 e 30 milioni di anni, con estesi fenomeni di silicizzazione epitermale associati a deposizione di minerali manganesiferi e ferruginosi. Interesse per oro epitermale offrono, sempre in Sardegna, le aree con vulcanismo pleistocenico e la zona corrispondente alla linea Sestri-Voltaggio, in Liguria.

Mineralizzazioni o impronte geochimiche che possano fare da indicatori dell'esistenza di depositi epitermali di metalli preziosi sono largamente presenti nel territorio italiano. In questo senso le aree più indiziate sono certamente quelle dell'alto Lazio-Toscana, dove notoriamente si ritrovano mineralizzazioni a mercurio, antimonio, piombo, zinco, argento, fluorite, nonché ad alunite, tutte strettamente associate e ottime indicatrici della presenza di metalli preziosi.

La presenza di porosità, sia primaria sia secondaria, è un fattore chiave da accertare nelle aree scelte sulla base di tutte le altre considerazioni sopramenzionate. Le aree in cui tale caratteristica è assente possono essere scartate senza indugio, in quanto prive di qualsiasi potenziale minerario. La valutazione di questo parametro può essere fatta in dettaglio sulla base di analisi stratigrafiche e strutturali, in quanto in fase riconoscitiva le sole informazioni che si possono avere sulla porosità sia primaria sia secondaria sono costituite dalla conoscenza della natura dei litotipi affioranti nell'area di indagine.

BIBLIOGRAFIA

LOWELL JAMES D. e GUILBERT JOHN M., *Lateral and Vertical Alteration-Mineralization Zoning in Porphyry Ore Deposits* in «Economic Geology», 65, pp. 373-408, 1970.

LEVINSON A. A., *Introduction to Exploration Geochemistry*, Applied Publications Ltd., Wilmette, Illinois, 1974.

ROSE ARTHUR W., HAWKES HERBERT E. e WEBB JOHN S., *Geochemistry in Mineral Exploration*, Academic Press, 1979.

HENLEY RICHARD W. ed ELLIS A. J., *Geothermal Systems Ancient and Modern, a Geochemical Review* in «Earth Sciences Review», 19, pp. 1-50, 1983.

PECCERILLO ANGELO e MANETTI PIERO, *The Potassium Alkaline Volcanism of Central-Southern Italy: a Review of the Data Relevant to Petrogenesis and Geodynamic Significance* in «Transactions of the Geological Society of South Africa», 88, pp. 379-394, 1985.

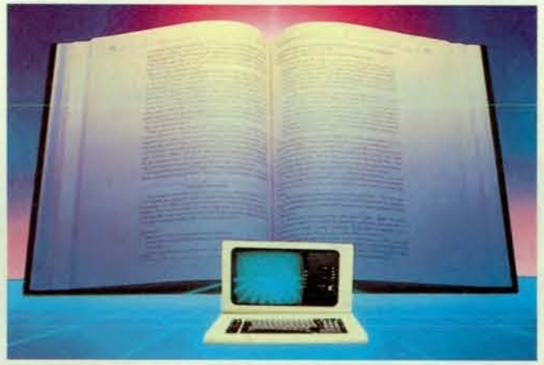
COX DENNIS P. e SINGER DONALD A., *Mineral Deposit Models*, United States Geological Survey Bulletin 1693, 1987.

ItalCable
Telecomunicazioni Internazionali

M.A.G.I.C. ON LINE

M.A.G.I.C. ON LINE con ITALCABLE

Anche Voi avete certamente bisogno di un contratto ed un accesso **unici** per tutte le banche dati con cui colloquiare. L'accesso in tempo reale ad informazioni specializzate è infatti la condizione essenziale per essere leader nella vostra attività.
Vi è indispensabile però un "sistema" capace di farVi navigare fino alla fonte meglio informata, con una procedura semplice ed in grado di offrire la consultazione più vasta possibile e senza complicazioni di carattere amministrativo e burocratico.



M.A.G.I.C. ON LINE È LA RISPOSTA

- con un **unico e semplice linguaggio** accedete ad oltre 1000 banche dati in tutto il mondo rese disponibili dai più importanti distributori internazionali.
- con un **unico contratto in Italia** siete abbonati ai servizi di oltre 1000 banche dati in tutto il mondo, risparmiando tempo e denaro.

ACCESSO AL SERVIZIO

- L'accesso al servizio è assicurato tramite la rete pubblica a commutazione di pacchetto ITAPAC. L'Italcable fornirà a tutti i clienti che ne faranno richiesta, un codice d'utente unico e valido per tutto il territorio nazionale.
- Il servizio è disponibile a 300 Baud su tutto il territorio nazionale e in alcune città a 1200 Baud.
- Per collegarsi al servizio **M.A.G.I.C. ON LINE** basta essere provvisti di:
  un terminale adatto (al terminale video o stampante TTY compatibile, personal computer con interfaccia di trasmissione dati asincrona, terminale telematico multifunzione);
  un modem o un accoppiatore acustico (se non sono integrati nel terminale);
  una linea telefonica
- Per facilitare coloro che non sono dotati di terminale il personale Italcable è a disposizione per il necessario supporto informativo.

Per ulteriori informazioni chiamare il Numero Verde Italcable (Lun.-Ven. 08,00-16,00)

CHIAMATA GRATUITA
NUMEROVERDE
1678 - 60025

Gruppo **IRI-STET**

Che cosa causò l'estinzione in massa?

Sessantacinque milioni di anni fa la Terra subì un improvviso mutamento. I rettili che avevano dominato il paesaggio scomparvero insieme a più di metà di tutte le specie di piante e animali terrestri e marini. Dal momento che una specie può sopravvivere anche quando la sua popolazione è fortemente ridotta, il fatto che tante specie si siano estinte alla fine del Cretaceo, al cosiddetto limite KT, fa pensare che per un certo periodo la Terra sia divenuta estremamente inospitale.

I mammiferi riuscirono in qualche modo a superare questi tempi duri. Occupando le nicchie ecologiche lasciate libere dai dinosauri assunsero una posizione di dominio fra gli animali di grossa taglia. Tra i discendenti di questi, una particolarissima specie avrebbe poi finito col mettersi a studiare la documentazione fossile delle proprie lontane origini.

Una documentazione costituita da tracce di ogni genere: oltre alle ossa di dinosauri, relativamente rare, vi sono infatti resti di molti altri organismi, inclusi abbondanti microfossili di foraminiferi e di altri organismi marini. Ma anche lo spessore e la composizione chimica degli strati hanno qualcosa da raccontare.

Tuttavia la storia cambia a seconda di chi la legge. Walter Alvarez e Frank Asaro dell'Università della California a Berkeley vedono i segni di un immane impatto da parte di un corpo celeste in alcuni strati di rocce sedimentarie dell'Appennino umbro e di altre località della Terra. Vincent E. Courtillot dell'Istitut de Physique du Globe di Parigi scorge la testimonianza di un mondo devastato e «inquinato» da grandiose eruzioni vulcaniche che hanno formato i trappi del Deccan, in India. Nelle pagine seguenti questi autori espongono le loro rispettive tesi. Nel frattempo i grandi rettili continuano a custodire il mistero della loro scomparsa.

Un impatto extraterrestre

Sono sempre più consistenti le prove che attribuiscono all'impatto di un corpo celeste le estinzioni in massa della fine del Cretaceo

di Walter Alvarez e Frank Asaro

Circa 65 milioni di anni fa qualcosa distrusse metà di tutte le forme di vita presenti sulla Terra, spazzando via i dinosauri, fino ad allora indiscussi dominatori del regno animale, e lasciando agli umili mammiferi l'eredità del vasto territorio che prima era stato di questi grandi rettili. Gli esseri umani, discendenti da questi superstiti, non possono non domandarsi chi, o che cosa, abbia provocato questa ecatombe e in che modo i mammiferi nostri lontani antenati siano riusciti a sopravvivere.

In quest'ultimo decennio ricercatori di tutto il mondo, in discipline che vanno dalla paleontologia all'astrofisica, hanno riunito le loro capacità di osservazione, di intuito sperimentale e di immaginazione teorica, per cercare di rispondere a queste domande. Si è passati attraverso lunghi mesi di meticolose misurazioni, periodi di confusione, lampi d'intuito e momenti di grande eccitazione, fino a quando i tasselli del mosaico hanno trovato la loro collocazione.

Oggi riteniamo di aver risolto il mistero. Circa 65 milioni di anni fa un gigantesco asteroide o una cometa piombò sulla Terra a una velocità di oltre 10 chilometri al secondo. L'enorme energia liberata da quell'impatto scatenò una spaventosa serie di catastrofi ambientali: tempeste, maremoti, freddo e buio ai quali seguirono riscaldamento per effetto serra, piogge acide e vastissimi incendi. Quando ritornò la quiete, metà della flora e della fauna si era estinta. La storia della Terra aveva imboccato una via del tutto nuova.

Per altri indiziati chiamati in causa per il «genocidio» dei dinosauri (per esempio cambiamenti del livello del mare, mutamenti climatici ed eruzioni vulcaniche), esistono «alibi» che ne consigliano l'esclusione. Tuttavia alcune questioni non sono ancora state chiarite: dove ebbe luogo l'impatto? Si trattò di un solo impatto o di più d'uno? Tali impatti avvennero a intervalli di tempo regolari? E quale sarebbe stato il ruolo di queste catastrofi nell'evoluzione?

L'enigma presentato da questa estinzione in massa può essere affrontato analogamente a come oggi si conducono le indagini su un delitto. Esistono prove - anomalie chimiche, granuli di minerali e rapporti isotopici (in luogo di sangue o impronte digitali o bustine di fiammiferi) - sparse un po' in tutto il mondo. Tuttavia non è rimasto alcun testimone oculare né vi è alcuna possibilità di strappare confessioni. Il trascorrere di milioni di anni ha distrutto o degradato gran parte delle prove relative al caso, lasciando solo gli indizi più deboli.

In effetti, è difficile persino essere sicuri che i singoli fossili a noi pervenuti siano proprio i resti delle vittime del catastrofico impatto. Ma i paleontologi sono certi che ci devono essere state delle vittime, visto che le rocce sedimentarie fossilifere mostrano una netta discontinuità 65 milioni di anni fa. Creature come dinosauri e ammoniti, abbondanti per decine di milioni di anni, scomparvero improvvisamente e per sempre. Molti altri gruppi di animali e di piante subirono una decimazione.

Questa discontinuità definisce il limite tra il Cretaceo (*Kreide* in tedesco) durante il quale i dinosauri regnarono incontrastati e il Terziario che vide l'affermarsi dei mammiferi, ed è nota come limite KT.

Quando cominciammo a studiare il limite KT, volevamo scoprire con precisione in che arco di tempo fosse avvenuta l'estinzione. Era stata improvvisa, ossia era durata pochi anni o secoli, o si era trattato di un evento graduale durato milioni di anni? Gran parte dei geologi e dei paleontologi aveva da sempre supposto che l'estinzione fosse stata lenta. (Queste discipline tendono per tradizione al gradualismo e non sono solite chiamare in causa catastrofi.) Dal momento che i resti di dinosauri sono relativamente rari, la loro età fornisce scarse informazioni sulla durata degli eventi che portarono all'estinzione. Si era comunque propensi a considerare l'estinzione dei dinosauri come avvenuta in maniera graduale.

Tuttavia i paleontologi, esaminando resti fossili di pollini o di animali marini unicellulari, come i foraminiferi, appurarono che l'estinzione era stata improvvisa. Questa prova era secondo noi più valida in quanto gli organismi di dimensioni più piccole lasciano una maggiore quantità di fossili permettendo di meglio definire il quadro temporale.

L'estinzione appare ancora più improvvisa quando si studiano i resti fossili di animali di media taglia, come gli invertebrati marini. Tra questi ci sono le ammoniti (parenti dell'attuale nautilo a conchiglia concamerata) estintesi alla fine del Cretaceo. La testimonianza migliore della loro estinzione si trova negli affioramenti costieri del Golfo di Biscaglia, al confine tra Spagna e Francia.

Nel 1986 Peter L. Ward e colleghi dell'Università di Washington hanno condotto studi dettagliati su questi affioramenti a Zumaya, in Spagna. Ward ha notato che le ammoniti sembravano essersi estinte gradualmente, scomparendo una specie dopo l'altra in un intervallo stratigrafico di circa 170 metri, che rappresenta pressappoco cinque milioni di anni. Ma nel 1988 lo stesso Ward ha studiato in Francia due sezioni vicine, trovando la prova che queste specie di ammoniti erano in realtà sopravvissute fino al limite KT. A Zumaya l'apparente gradualità dell'estinzione era semplicemente dovuta all'incompletezza della documentazione fossile. Se gli organismi i cui resti fossili sono ben conservati si estinsero improvvisamente, allora è verosimile che anche gli altri animali, come i dinosauri, i cui resti sono più scarsi abbiano subito la stessa sorte.

Ciò stabilisce dunque che l'estinzione sia stata improvvisa (in termini geologici), ma non precisa in quanti anni sia avvenuta: è infatti arduo stabilire la datazione assoluta di una roccia con una precisione di un milione di anni ed anche una datazione relativa basata sulla stratificazione geologica è sempre approssimata di 10 000 anni, un periodo più lungo dell'intero lasso di tempo in cui si è sviluppata la civiltà umana.

Tuttavia si può stimare la durata dell'estinzione in massa che segna il limite KT in modo più preciso. A Gubbio, in Italia, nei calcari sedimentati in acque profonde, un sottile strato di argilla separa i sedimenti del Cretaceo da quelli del Terziario. Lo straterello, scoperto da Isabella Premoli Silva dell'Università di Milano, è spesso circa un centimetro. Negli anni settanta uno di noi (Alvarez) aveva fatto parte di un gruppo che aveva individuato questi depositi di argilla in calcari di sei metri di spessore, sedimentati durante un periodo di polarità geomagnetica invertita (designato con la sigla 29R) durato 500 000 anni. A prima vista si potrebbe valutare che lo strato d'argilla, e l'estinzione in massa da esso contrassegnata, corrisponda a un arco di tempo di un migliaio di anni.

Jan Smit dell'Università di Amsterdam ha condotto uno studio analogo sui sedimenti di Caravaca, nella Spagna meridionale, dove la stratigrafia è anche più precisa, valutando l'arco di tempo dell'estinzione in non più di 50 anni. Nella scala dei tempi geologici è un evento accaduto con rapidità spaventosa.

Il nostro lavoro sul limite KT cominciò verso la fine degli anni settanta quando noi e i nostri colleghi di Berkeley, Luis W. Alvarez (il noto fisico, padre di Walter) ed Helen V. Michel, cercammo di sviluppare un metodo più accurato per determinare in quanto tempo si fosse depositato lo strato di argilla del limite KT di Gubbio. I nostri sforzi fallirono, ma condussero a una prima prova decisiva sull'identità del «colpevole». (Questo è ciò di cui gli scienziati, come gli investigatori, hanno bisogno: un duro lavoro e un colpo di fortuna che, nel nostro caso, era la scoperta dello straterello di argilla.)

Il metodo si basava sulla rarità dell'iridio nella crosta terrestre (circa 0,03 parti per miliardo) in confronto con le concentrazioni di 500 parti per miliardo delle antiche meteoriti litoidi (condriti carbonacee). L'iridio terrestre è raro nella crosta, perché per la maggior parte è legato al ferro del nucleo.

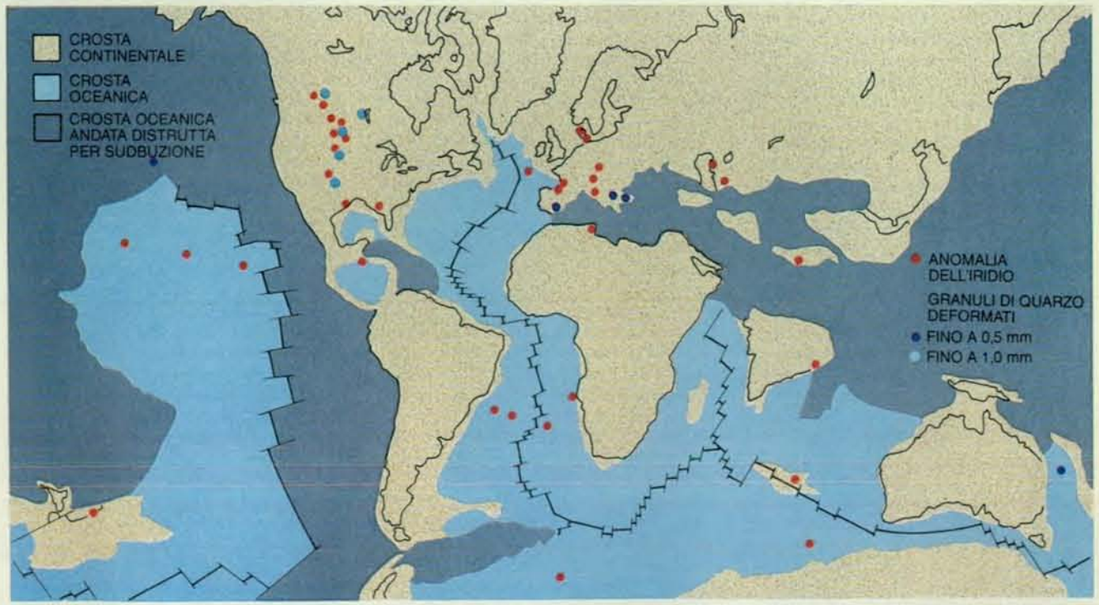
Abbiamo pensato che l'iridio fosse finito nei sedimenti di mare profondo, come quelli di Gubbio, principalmente attraverso le continue piogge di micrometeoriti, la cosiddetta «polvere cosmica». Questa caduta costante potrebbe dare una misura del tempo: più è abbondante l'iridio contenuto in uno strato sedimentario, più a lungo deve essere durata la sua deposizione. Inoltre, l'iridio potrebbe venire misurato anche a concentrazioni molto basse con l'analisi per attivazione neutronica, una tecnica basata sul bombardamento di neutroni che rende il metallo radioattivo e quindi rilevabile.

L'ipotesi che abbiamo preso in considerazione era che lo strato di argilla al limite KT si fosse formato in circa 10 000 anni, in assenza ormai di organismi a guscio calcareo e quindi senza che avvenisse deposizione di carbonato di calcio (che costituisce gran parte del calcare). A Gubbio, la maggior parte degli strati contiene circa il 95 per cento di carbonato di calcio e il 5 per cento di argilla, mentre lo strato limite contiene il 50 per cento di argilla. Se questa ipotesi fosse corretta, il rapporto iridio/argilla nello straterello limite e negli strati rispettivamente superiore e inferiore dovrebbe essere simile. Se invece il contenuto di argilla fosse diminuito contemporaneamente al carbonato di calcio, nello strato limite si avrebbe un rapporto iridio/argilla più alto rispetto a quello delle rocce adiacenti.

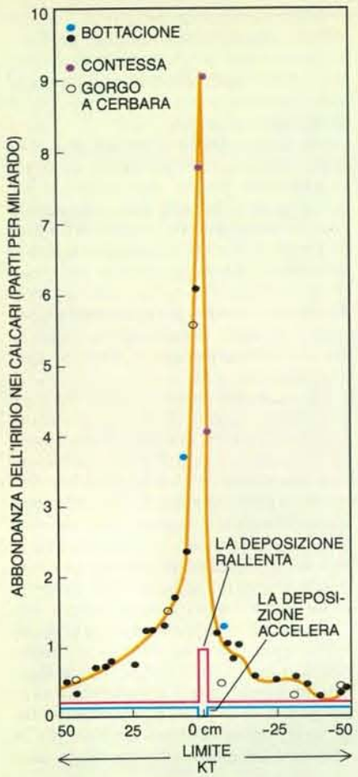
Nel giugno del 1978 furono pronte le nostre prime analisi sull'iridio rinvenuto a Gubbio. Potete immaginare il nostro stupore e il nostro disorientamento quando vedemmo che l'argilla dello strato limite e il calcare immediatamente adiacente contenevano molto più iridio che non in qualsiasi nostra ipotesi precedente, una quantità paragonabile a quella che in tutto il resto della roccia si era depositata durante i 500 000 anni dell'intervallo 29R.

Chiaramente questa concentrazione non poteva essere attribuita alla solita pioggia di polvere cosmica. Per un anno discutemmo le possibili fonti, esaminando e scartando un'idea dopo l'altra. Poi nel 1979 proponemmo la sola soluzione sopravvissuta alle nostre verifiche: una grossa cometa o un asteroide di circa 10 chilometri di diametro doveva aver colpito la Terra, immettendo nell'atmosfera un'enorme quantità di iridio.

Da quando proponemmo per la prima volta l'ipotesi dell'impatto sono emerse ulteriori prove che l'hanno resa sempre più verosimile. Più di 100 scienziati in 21 laboratori di 13 paesi hanno trovato negli strati corrispondenti al limite KT livelli di iridio anormalmente alti in circa



Le prove del catastrofico impatto che causò le estinzioni in massa alla fine del Cretaceo sono distribuite su tutto il globo. Questa carta mostra la posizione dei continenti 65 milioni di anni fa. Livelli anomali di iridio - prova quasi certa di un impatto di enormi proporzioni - sono stati trovati in sedimenti al limite Cretaceo-Terziario (KT) insieme a cristalli di quarzo con tipiche strutture da shock. Non è però mai stato trovato il cratere del diametro di circa 150 chilometri che un impatto del genere avrebbe dovuto produrre.



L'anomalia nel contenuto di iridio misurata in Umbria nei sedimenti calcarei che segnano il limite KT è un chiaro indizio dell'impatto di un corpo extraterrestre. L'iridio infatti è raro nella crosta terrestre e solo un corpo celeste potrebbe aver depositato la quantità rinvenuta. Sono state proposte altre ipotesi per spiegare l'estinzione in massa (*linee in rosso e blu*) nelle quali viene fatta variare la velocità di deposizione del calcare tenendo costante la caduta di iridio cosmico, ma queste ipotesi non riescono a giustificare contenuti di iridio così elevati.

95 località sparse in tutto il mondo. L'anomalia è stata trovata in sedimenti marini e non marini, in affioramenti continentali e in carote di sedimenti oceanici. Questo fatto è particolarmente significativo in quanto, dopo aver analizzato una grandissima quantità di sedimenti, abbiamo appurato che le anomalie nel contenuto di iridio sono di solito molto rare. Anzi, per quanto ne sappiamo, quella del limite KT è praticamente unica.

L'anomalia dell'iridio si spiega bene con l'ipotesi dell'impatto dal momento che, nello strato limite, la quantità di iridio rispetto agli elementi con comportamento chimico simile, come platino, osmio, rutenio, rodio e oro è la stessa di quella contenuta nelle meteoriti. Miriam Kastner della Scripps Institution of Oceanography, che fa parte del nostro gruppo di lavoro, ha stabilito che il rapporto oro/iridio del limite KT, accuratamente studiato a Stevns Klint in Danimarca, è in accordo, entro il 5 per cento di errore, con quello della maggior parte delle antiche meteoriti litoidi (condriti carbonacee di I tipo).

In realtà, i rapporti di tutti gli elementi del gruppo del platino trovati al limite KT sono a favore della causa extraterrestre. George Bekov dell'Istituto di spettroscopia di Mosca e uno di noi (Asaro) hanno trovato che l'abbondanza relativa di rutenio, rodio e iridio può distinguere le meteoriti litoidi dai campioni terrestri. L'analisi dei campioni del limite KT provenienti da Stevns Klint, dal Turkmenistan (URSS) e da altre località avvalorano l'ipotesi dell'impatto.

Lo stesso vale per i rapporti tra gli isotopi. Jean-Marc Luck dell'Institut de Physique du Globe di Parigi e Karl K. Turekian della Yale University hanno trovato che gran parte dell'osmio nei campioni del limite KT provenienti dalla Danimarca e dal New Mexico non può avere origine continentale avendo un contenuto troppo basso di osmio 187. Il rapporto tra osmio 187 e osmio 186 è più alto nelle rocce continentali che nelle meteoriti o nel mantello terrestre, perché queste rocce sono relativamente ricche di renio, il cui isotopo radioattivo, il renio 187, decade a osmio 187. L'osmio nei campioni del limite KT deve avere origine extraterrestre o deve provenire dal mantello terrestre.

Non solo la composizione delle rocce al limite KT fa pensare all'impatto, ma pure la loro mineralogia. Nel 1981 Smit ha scoperto un altro indizio significativo: sferule di minerali del diametro di un millimetro nell'argilla KT a Caravaca. (Alessandro Montanari di Berkeley ne ha confermato la presenza anche in Italia.) Le sferule si sarebbero originate da goccioline di roccia basaltica fusa dall'impatto e raffreddatasi rapidamente durante la traiettoria balistica all'esterno dell'atmosfera, quindi alterata chimicamente e trasformata nell'argilla dello strato limite. Esse sono l'equivalente basaltico delle tectiti vetrose, più ricche di silicati, e delle microtectiti, che sono l'effetto di impatti più piccoli. Il chimismo basaltico fa pensare che l'impatto sia avvenuto su crosta oceanica.

Oltre alle sferule, Bruce F. Bohor dello US Geological Survey di Denver e Donald M. Triplehorn dell'Università dell'Alaska hanno scoperto granuli di quarzo con strutture da impatto. Gli accurati studi di E. E. Foord, Peter J. Modreski e Glen A. Izett dell'USGS rivelano che i granuli presentano una serie di lamelle planari che si intersecano (bande di deformazione), indicatrici di shock da ipervelocità. Questi granuli si trovano solo nei crateri da impatto conosciuti, nei siti di esperimenti nucleari, in materiali sottoposti in laboratorio a shock estremi e, appunto, nel limite KT.

C'è in effetti a Manson nell'Iowa, sotto il detrito glaciale, un cratere che fa al caso nostro; si trova in una roccia in posto ricca di quarzo e la sua posizione è adatta a spiegare la dimensione e il gran numero dei granuli di quarzo da impatto. Con i suoi 32 chilometri di diametro, il cratere è troppo piccolo per essere stato formato da quel singolo corpo celeste ritenuto responsabile dell'estinzione. Nondimeno, studi dettagliati del cratere rivelano che esso ha un'età non distinguibile da quella del limite KT, cosicché anch'esso deve avere svolto un ruolo nell'evento misterioso su cui stiamo indagando.

Come potrebbe un solo impatto disperdere materiali in tutto il globo? Un asteroide di 10 chilometri che si muovesse a più di 10 chilometri al secondo provocherebbe nell'atmosfera un enorme vuoto d'aria. All'impatto con il suolo l'energia cinetica dell'asteroide si trasformerebbe in calore, in una esplosione non nucleare 10 000 volte più potente di quella che potrebbe provocare l'intero arsenale mondiale di armi nucleari. Parte del materiale dell'asteroide e della roccia terrestre in prossimità del punto d'impatto verrebbe vaporizzata e sfuggirebbe attraverso il vuoto d'aria, prima che questo abbia il tempo di ricolmarsi.

La sfera di fuoco formata dal gas incandescente prodotto dall'esplosione espellerebbe il materiale dall'atmosfera. La sfera di fuoco prodotta da un'esplosione nucleare atmosferica si espande fino a raggiungere la stessa pressione dell'atmosfera circostante, quindi si innalza fino a raggiungere la quota in cui la sua densità uguaglia quella dell'aria. A questo punto, a circa 10 chilometri di altezza, il gas si espande lateralmente a formare l'ombrello della nota nube a fungo.

Modelli al calcolatore di esplosioni con energie di 1000 megaton (20 volte l'energia delle più potenti bombe nucleari, ma solo 1/100 000 dell'energia dell'impatto KT) hanno dimostrato che la sfera di fuoco non raggiunge mai una pressione uguale a quella dell'atmosfera circostante. Invece, quando la sfera di fuoco giunge a quote alle quali la densità dell'atmosfera diminuisce significativamente, la sua ascesa è più rapida e il gas abbandona la parte alta dell'atmosfera a una velocità sufficiente a sfuggire al campo gravitazionale terrestre. La sfera di fuoco conseguente all'impatto di un asteroide ancora più grande supererebbe probabilmente il limite dell'atmosfera, trasportando tutto il materiale coinvolto nell'impatto e scagliandolo in orbite da cui esso potrebbe ricadere in qualsiasi regione della Terra.

L'impatto di un corpo delle dimensioni di una cometa produrrebbe sulla Terra un cratere di 150 chilometri di diametro e annienterebbe sicuramente ogni forma di vita nel raggio della sfera di fuoco. Si sta affinando nei modelli la comprensione dei fenomeni che potrebbero seguire un impatto in grado di provocare estinzioni su scala globale. Si tratterebbe di un misto di buio, freddo, fuoco, piogge acide e, successivamente, riscaldamento per effetto serra.

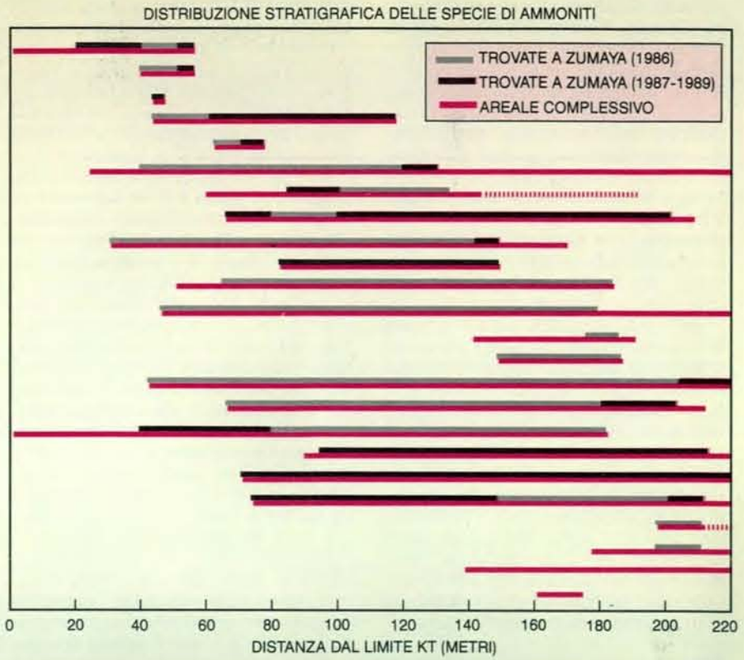
Nel nostro lavoro originale proponevamo che la polvere generata dall'impatto avesse causato un oscuramento totale. Secondo le simulazioni al calcolatore realizzate nel 1980 da Richard P. Turco dell'R&D Associates, O. Brian Toon della National Aeronautics and Space Administration e colleghi, la polvere immessa nell'atmosfera in seguito all'impatto di un corpo di 10 chilometri di diametro ostacolerebbe la luce a tal punto che per molti mesi non si riuscirebbe letteralmente a vedere a un palmo dal naso.

Senza la luce la fotosintesi delle piante si arresterebbe e le catene alimentari si interromperebbero ovunque. L'oscurità produrrebbe anche temperature estremamente basse, il cosiddetto «inverno da impatto». (Dopo aver considerato gli effetti dell'impatto, Turco, Toon e colleghi sono passati a studiare l'inverno nucleare, uno scenario altrettanto catastrofico che questi autori hanno descritto nell'articolo *Gli effetti sul clima di una guerra nucleare* in «Le Scienze» n.194, ottobre 1984.)

Nel 1981 Cesare Emilliani dell'Università di Miami, Eric Krause dell'Università del Colorado ed Eugene M. Shoemaker dell'USGS hanno indicato che un impatto oceanico immetterebbe nell'atmosfera non solo rocce polverizzate, ma anche vapore d'acqua. Il vapore, che è in grado di trattenere il calore emesso dalla Terra, rimarrebbe in quota molto più a lungo della polvere, cosicché l'inverno da impatto sarebbe seguito da un riscaldamento per effetto serra. In tempi più recenti John D. O'Keefe e Thomas J. Ahrens del California Institute of Technology hanno proposto che l'impatto possa essere avvenuto in un'area calcarea, con liberazione di grandi volumi di anidride carbonica, altro gas responsabile dell'effetto serra. Molte piante e animali sopravvissuti ai rigori dell'inverno da impatto potrebbero essere stati uccisi da un periodo successivo di calore eccezionale.

Nel frattempo John S. Lewis, G. Hampton Watkins, Hyman Hartman e Ronald G. Prinn del Massachusetts Institute of Technology hanno calcolato che il riscaldamento dell'atmosfera provocato dall'impatto stesso avrebbe innalzato la temperatura a un livello tale da far combinare l'ossigeno e l'azoto dell'aria. L'ossido di azoto avrebbe finito col precipitare come acido nitrico in una pioggia acida continua. Questo meccanismo può spiegare bene la diffusa estinzione di piante e di invertebrati marini, i cui gusci carbonatici sono solubili nelle acque acide.

Un altro meccanismo distruttivo è venuto alla luce grazie alla scoperta, da



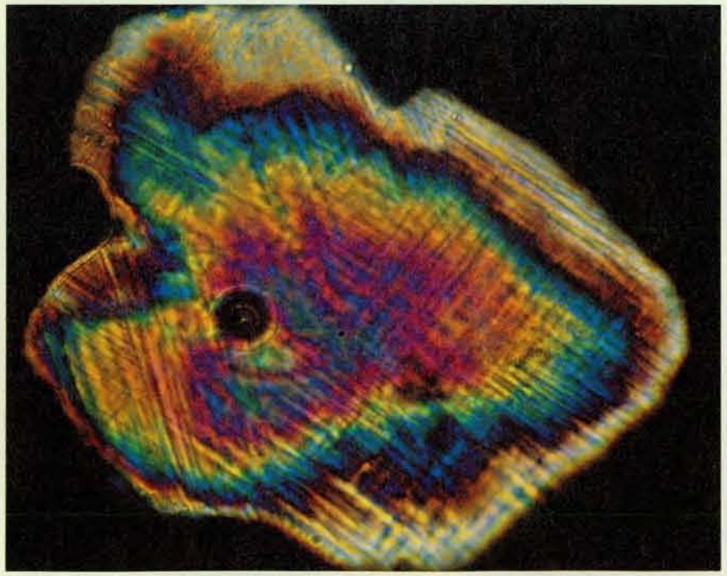
La distribuzione dei fossili marini evidenzia le numerose scomparse improvvise alla fine del Cretaceo. Sebbene lo studio di alcune rocce sembri deporre a favore della gradualità, la documentazione fossile completa indica che le specie si estinsero contemporaneamente.

In parecchie località sono state trovate incluse in argille del limite KT sferule basaltiche. Le sferule, ora alterate chimicamente, sono state originate dall'impatto come goccioline di materiale fuso della crosta oceanica che in seguito sono ricadute su tutto il globo.

parte di Wendy Wolbach, Ian Gilmore ed Edward Anders dell'Università di Chicago, di una gran quantità di fuliggine nell'argilla del limite KT. Se l'argilla si fosse depositata in pochi anni (o meno) la quantità di fuliggine al limite indicherebbe un improvviso incendio di vegetazione equivalente a metà delle foreste esistenti attualmente nel mondo. Jay Meos dell'Università dell'Arizona e colleghi hanno calcolato che la radiazione infrarossa dei materiali incandescenti espulsi dall'impatto potrebbe aver innescato, al loro rientro nell'atmosfera, incendi in tutto il globo.

Studi dettagliati dei sedimenti al limite KT possono in conclusione fornire prove a favore di un particolare meccanismo distruttivo. Per esempio, l'episodio di dissoluzione presente nei calcari italiani testimonia che subito dopo il livello di estinzione le acque più profonde erano acide subito dopo il limite che testimonia l'estinzione. E il lavoro che abbiamo fatto con William Lowrie dell'ETH di Zurigo dimostra anche che quelle acque cambiarono rapidamente dal loro normale stato ossidante a una condizione riducente, probabilmente a causa della morte in massa degli organismi marini.

Purtroppo nessuno ha mai potuto trovare un cratere di 150 chilometri, il prodotto dall'impatto di un corpo di 10 chilometri di diametro. Il cratere potrebbe essere nascosto sotto la coltre glaciale antartica o potrebbe essere stato su quel 20 per cento di superficie terrestre consumato per subduzione ai margini delle zolle oceaniche. Gli elementi che permetterebbero di individuarne la posizione sono contraddittori: le sferule basaltiche nell'argilla del limite indicano un impatto su fondo oceanico, ma i granuli di quarzo deformati testimonierebbero piuttosto un impatto su continente.



Questo granulo di quarzo mostra deformazioni laminari generate in seguito a shock da ipervelocità. Queste strutture del quarzo, che si ritrovano in siti di esplosioni nucleari e in prossimità di crateri da impatto meteoritico, sono state anche riprodotte in laboratorio.

Si è proposto recentemente che l'estinzione al limite KT sia stata causata da due o più impatti quasi simultanei. Shoemaker e Piet Hut dell'Institute for Advanced Study di Princeton, nel New Jersey, hanno identificato una serie di meccanismi che potrebbero produrre impatti multipli contemporanei o distribuiti nel corso di molti anni. Su Terra, Luna e altri pianeti si sono trovati crateri doppi e multipli, cosa che fa supporre che alcuni asteroidi possano essere costituiti da due o più corpi orbitanti reciprocamente l'uno intorno all'altro. In alternativa, la Terra potrebbe essere stata colpita da due o più grossi frammenti di un nucleo di cometa in disintegrazione.

Impatti multipli avvenuti in periodi di tempo più lunghi potrebbero essersi verificati se da un nucleo di cometa si fossero separati parecchi grossi frammenti che percorrevano orbite che incrociavano la Terra. Impatti di questo tipo potrebbero anche avvenire casualmente se qualche altro fattore aumentasse il numero medio di comete all'interno del sistema solare. Sebbene nessuna di queste ipotesi presa singolarmente sia convincente, nel loro insieme esse dimostrano che gli impatti multipli non sono così improbabili come si potrebbe pensare.

La teoria della cometa acquista credibilità per la scoperta di materiali apparentemente extraterrestri in prossimità del limite KT. Meixun Zhao e Jeffrey L. Bada dell'Università della California a San Diego hanno analizzato in Danimarca gli strati calcarei appena al di sopra e al di sotto del limite KT. Essi hanno trovato amminoacidi che non sono metabolizzati dagli organismi viventi sulla Terra, ma si trovano nelle condriti carbonacee. Sembra improbabile che gli amminoacidi possano resistere al calore di un impatto di grande entità e, in effetti, essi non compaiono nel limite KT stesso.

Kevin Zahnle e David Grinspoon della NASA hanno proposto che polvere proveniente da una cometa in disintegrazione sia entrata nell'atmosfera terrestre per un lungo periodo, trasportando con sé questi amminoacidi di origine extraterrestre. Durante quest'intervallo l'impatto di un grosso frammento della cometa avrebbe causato l'estinzione KT.

Una linea d'indagine apparentemente indipendente, basata su analisi statistiche piuttosto che chimiche, ha dato origine a una ipotesi che spiega come le comete potrebbero periodicamente colpire la Terra. Nel 1984 David M. Raup e John J. Sepkoski, Jr., dell'Università di Chicago hanno pubblicato uno studio sui resti fossili che sembra indicare che le estinzioni in massa si siano verificate a intervalli di 32 milioni di anni. Come gran parte degli scienziati che lavorano sul limite KT, eravamo molto scettici sui loro risultati, ma l'astrofisico Richard A. Muller dell'Università della California a Berkeley ha riesaminato i dati di Raup e Sepkoski e si è convinto che la periodicità da loro scoperta era reale.

Muller, Marc Davis di Berkeley e Hut hanno ipotizzato che una stella compagna poco luminosa e sconosciuta che compiva un'orbita intorno al Sole in 32 milioni di anni (denominata provvisoriamente Nemesis) avrebbe potuto con regolarità turbare le orbite di comete che viaggiavano ai confini del sistema solare. L'interferenza avrebbe inviato all'interno del sistema solare una tempesta di comete per un milione di anni, facendo aumentare molto la probabilità di un grosso impatto (o di molti) sulla Terra. Daniel Whitmire della Università della Southwestern Louisiana e Albert Jackson della Computer Sciences Corporation hanno indipendentemente proposto la stessa ipotesi.

Quando Muller mi mostrò [è Alvarez che racconta] il lavoro che proponeva Nemesis, fui molto scettico. Gli dissi che pensavo si trattasse dell'«ingegnosa soluzione a un non-problema», non essendo convinto della prova di Raup e Sepkoski per le estinzioni in massa periodiche. Se l'ipotesi fosse stata corretta - sostenevo - i crateri terrestri da impatto avrebbero mostrato la stessa periodicità. Muller e io scoprimmo più avanti, con

I crateri gemelli dei laghi Clearwater, in Canada, dimostrano che sulla Terra gli impatti multipli non sono improbabili. Si può ipotizzare che almeno due impatti quasi contemporanei si siano resi responsabili delle estinzioni in massa della fine del Cretaceo.

sua soddisfazione e mia sorpresa, che le età dei crateri mostrano fondamentalmente la stessa periodicità delle estinzioni in massa. Da allora mi convinsi che l'ipotesi dovesse almeno essere seriamente presa in considerazione.

Tuttavia risulta molto difficile individuare una stella rossa poco luminosa vicina al Sole quando non si ha idea di dove cercarla. Muller e Saul Perlmutter di Berkeley sono attualmente a metà strada nella ricerca con telescopio computerizzato di una stella con le caratteristiche di Nemesis; contano di finire in un paio d'anni. Nel frattempo una nuova analisi delle età dei crateri e di quelle delle estinzioni in massa ha sollevato domande sull'effettiva periodicità del fenomeno. Il piccolo numero di eventi e la lacunosa informazione disponibile rendono difficile una risposta certa.

In un'indagine su un assassinio gli indiziati devono avere mezzi, movente e occasione. Un impatto avrebbe avuto senza dubbio i mezzi per causare l'estinzione di massa alla fine del Cretaceo e il fatto che si sia verificato un impatto proprio al momento giusto indica l'occasione. L'ipotesi dell'impatto fornisce, se non il movente, almeno la dinamica del crimine. Ma come stanno le cose per quanto riguarda altri eventuali indiziati?

Alcuni hanno un alibi inattaccabile: non potrebbero infatti aver ucciso una così grande varietà di forme viventi. Sostenere che i mammiferi abbiano predato le uova dei dinosauri - ipotesi assai datata - non potrebbe comunque spiegare la contemporanea estinzione di foraminiferi e ammoniti.

Stefan Gartner della Texas A&M University in passato ha ipotizzato che la vita marina fosse stata annientata da un immane e improvviso afflusso di acqua dolce dall'Oceano Artico che, rimasto sicuramente isolato dagli altri oceani durante il tardo Cretaceo, si doveva essere appunto riempito di acqua dolce. Ma questo meccanismo non può spiegare l'estinzione dei dinosauri o la scomparsa di molte specie di piante terrestri.

Vi potrebbero essere altri indiziati, ma tutti posseggono alibi temporali. Alcuni ricercatori hanno tentato di attribuire la responsabilità dell'estinzione in massa ad altri fattori come i cambiamenti del clima o del livello del mare. Tuttavia questi eventi hanno solitamente luogo in tempi molto più lunghi di quanto sia avvenuto in questo caso; inoltre non sembra che eventi di questo genere siano coincisi con l'estinzione, e comunque fatti analoghi si sono verificati più volte nella storia della Terra senza essere accompagnati da estinzioni.

Altri considerano il vulcanismo tra i principali indiziati. La prova più convincente che chiama in causa i vulcani si trova in India ed è un'enorme effusione di lava basaltica, i trappi del Deccan (Deccan Traps), avvenuta circa 65 milioni di anni fa. Un recente lavoro paleomagnetico di Vincent E. Courtillot (si veda l'articolo *Un'eruzione vulcanica* a pagina 47) e colleghi di Parigi conferma gli studi precedenti. Egli afferma che gran parte dei trappi del Deccan fu effusa entro un periodo di polarità geomagnetica invertita, con scarse sovrapposizioni nei periodi di polarità normale precedente e successivo. Il gruppo di Parigi ha trovato che l'intervallo in questione è probabilmente il 29R, durante il quale avvenne l'estinzione KT, anche se potrebbe essere pure l'intervallo di polarità invertita immediatamente precedente o successivo al 29R.

Dal momento che l'effusione che formò i trappi del Deccan è cominciata in un intervallo normale ed è terminata nel successivo, le eruzioni devono essere avvenute in almeno 500 000 anni. Perciò gran parte degli studiosi non ha considerato il vulcanismo un serio «indiziato» dello sterminio, che con ogni probabilità avvenne in 1000 anni o meno.

Alcuni ricercatori sostengono che, a dispetto della documentazione fossile, le estinzioni del limite KT si sarebbero svolte in molte migliaia di anni e che il vulcanismo potrebbe spiegare i granuli

di quarzo deformati, le sferule e anche l'anomalia dell'iridio.

Nel 1983 William H. Zoller e colleghi dell'Università del Maryland a College Park hanno trovato alte concentrazioni di iridio nei vapori del vulcano Kilauea nell'isola di Hawaii, raccolti su filtri posti a 50 chilometri di distanza; tuttavia il rapporto tra iridio e altri elementi rari nei vapori vulcanici non corrisponde al rapporto trovato nello strato corrispondente al limite KT. Il rapporto tra oro e iridio nel vapore del Kilauea è oltre 35 volte più alto di quello misurato al limite KT a Stevns Klint.

Ci si è anche chiesti se un'eruzione vulcanica esplosiva sia o meno in grado di produrre quarzo con struttura da shock. Tuttavia sembra ora assodato che le esplosioni vulcaniche possono produrre qualche deformazione, ma che le caratteristiche lamelle multiple che si riscontrano nel quarzo del limite KT possono formarsi solo per shock da impatto. Inoltre, John McHone dell'Arizona State University ha scoperto che queste lamelle contengono stishovite, una forma di quarzo che si genera solo a pressioni assai più elevate di quelle delle eruzioni vulcaniche. Mark H. Anders della Columbia University e Michael R. Owen della St. Lawrence University hanno usato una tecnica, la catodoluminescenza, con la quale si porta il quarzo alla luminescenza, per determinarne l'origine. I colori prodotti dai granuli non si accordano con l'origine vulcanica, ma indicano piuttosto un impatto avvenuto su una comune arenaria.

Per di più, le sferule basaltiche del limite KT testimoniano in ogni caso a sfavore del vulcanismo esplosivo: le sferule potrebbero essere state generate da forme più tranquille di vulcanismo, che certamente non avrebbe potuto distribuirle in tutto il mondo.

La chiara distribuzione globale dell'anomalia dell'iridio, del quarzo con struttura da shock e delle sferule basaltiche è una prova significativa che esclude il vulcanismo e dà sostegno all'ipotesi dell'impatto. Le eruzioni avvengono alla base dell'atmosfera: nel migliore dei casi inviano materiale nell'alta stratosfera. Se le sferule e i granuli di quarzo provenissero da un'eruzione, sarebbero rapidamente rallentati dall'attrito con l'atmosfera e ricadrebbero al suolo.

Nondimeno, le enormi eruzioni che crearono i trappi del Deccan sono avvenute durante un periodo che comprende l'estinzione contrassegnata dal limite KT. Inoltre rappresentano la più grande effusione di lava sulla Terra degli ultimi duecentocinquanta milioni di anni (sebbene volumi maggiori continuino a fuoriuscire delle dorsali medio-oceaniche). Nessun ricercatore può ignorare questo tipo di coincidenza.

Pare possibile che un impatto abbia innescato il vulcanismo dei trappi del Deccan. Se un corpo di enormi dimensioni colpisse la Terra, il cratere iniziale sarebbe profondo 40 chilometri e il venir meno della pressione farebbe fondere le rocce del sottostante mantello. Tuttavia anche i massimi vulcanologi trovano molto arduo spiegare in dettaglio come un impatto potrebbe innescare effusioni basaltiche su vasta scala.

Negli ultimi anni il dibattito tra i sostenitori di ciascuna ipotesi si è polarizzato: i fautori dell'impatto sono inclini a considerare i trappi del Deccan come irrilevanti, mentre i sostenitori del vulcanismo cercano di sminuire le prove dell'impatto, affermandone la compatibilità con il vulcanismo. Si va a nostro avviso su schemi di tipo hegeliano: si pongono una tesi (l'impatto) e un'antitesi (il vulcanismo) in cerca di una sintesi ancora tutta da definire.

Pur in questa incertezza di fondo l'evento misterioso che causò l'estinzione in massa ci offre numerose lezioni. Alla fine del diciottesimo secolo, mentre lo studio della Terra si avviava a diventare scienza, si accesero controversie destinate a durare tra i catastrofisti, i quali sostenevano che per l'evoluzione del pianeta fossero stati decisivi grandi eventi improvvisi, e gli attualisti, che spiegavano tutta la storia della Terra in termini di cambiamento graduale. Gli attualisti trionfarono a tal punto in questa disputa che per generazioni intere di geologi qualunque tipo di catastrofismo sarebbe stato bollato come non scientifico. Tuttavia, nell'universo domina la violenza, come l'astronomia ha dimostrato, e sta diventando finalmente chiaro che anche la storia della Terra ha avuto i suoi episodi violenti.

La prova che un gigantesco impatto fu responsabile delle estinzioni alla fine del Cretaceo ha definitivamente reso degno di rispetto il punto di vista dei catastrofisti. In futuro, geologi liberi di pensare sia in termini attualisti che catastrofisti avranno migliori possibilità dei loro predecessori di comprendere appieno i processi e la storia del pianeta.

Le catastrofi hanno un ruolo importante anche nel pensiero evoluzionistico. Se un impatto fortuito avvenuto 65 milioni di anni fa può aver distrutto metà delle forme di vita, allora la selezione naturale non è il solo fattore trainante dell'evoluzione. Le specie non solo devono essere ben adattate, ma devono anche avere fortuna.

Se disastri casuali possono episodicamente annientare intere schiere di organismi ben adattati, allora la storia della vita non è prestabilita così come non è detto che l'evoluzione tenda inesorabilmete a dare origine a forme di vita intelligente come gli esseri umani. Anzi, Norman Sleep della Stanford University e colleghi hanno proposto che nelle prime fasi della storia terrestre, quando gli impatti di asteroidi erano più frequenti, la vita ai suoi primordi possa essersi estinta più di una volta.

Le catastrofi da impatto possono anche impedire che il corso dell'evoluzione si areni. I resti fossili indicano che in tempi normali ciascuna specie sviluppa un adattamento progressivamente superiore alla propria nicchia ecologica, per cui diventa sempre più difficile per altre specie svilupparsi entro quella stessa nicchia, un fenomeno che fa rallentare la velocità dell'evoluzione. A questo punto potrebbe intervenire la soppressione in massa di specie in seguito a un impatto che darebbe alle specie superstiti la grande opportunità di evolversi in nicchie rimaste sguarnite. (Abbiamo sentito non pochi studenti fantasticare per analogia sulle eccellenti prospettive di lavoro che per loro si schiuderebbero se metà di tutti i professori di ruolo fosse improvvisamente licenziata.) In effetti, la documentazione fossile mostra un aumento della velocità di evoluzione subito dopo la fine del Cretaceo.

Tra i fortunati superstiti dell'estinzione KT vi furono i primi mammiferi del Terziario, nostri antenati. Quando i dinosauri dominavano sulla Terra, pare che i mammiferi fossero piccoli e insignificanti. Il metabolismo a sangue caldo, la piccola taglia, il loro grande numero o altre caratteristiche possono averli messi in grado di superare le dure condizioni successive all'impatto. E con l'uscita di scena dei grandi rettili, i mammiferi cominciarono quella fase esplosiva dell'evoluzione che finì col produrre l'intelligenza umana. Chi scrive si trova a riflettere di tanto in tanto sul fatto che dobbiamo la nostra esistenza di esseri pensanti a quel catastrofico impatto che 65 milioni di anni fa portò all'estinzione i dinosauri.

BIBLIOGRAFIA

ALVAREZ LUIS W., ALVAREZ WALTER, ASARO FRANK e MICHEL HELEN V., *Extraterrestrial Cause for the Cretaceous--Tertiary Extinction* in «Science», 208, n. 4448, 6 giugno 1980.

SILVER LEON T. e SCHULTZ PETER H. (a cura), *Geological Implications of Impacts of Large Asteroids and Comets on the Earth*, in «Geological Society of America Special Paper», 190, 1982.

ALVAREZ W., *Toward a Theory of Impact Crises* in «Eos», 67, n. 35, 2 settembre 1986.

RAUP DAVID M., *Nemesis Affair: A Story of the Death of Dinosaurs and the Ways of Science*, W. W. Norton & Co., 1986.

MULLER R. A., *Nemesis*, Weidenfeld and Nicolson, 1988.

SHARPTON V. L. e WARD P. D. (a cura), *Global Catastrophes in Earth History: An Interdisciplinary Conference on Impacts, Volcanism, and Mass Mortality*, in «Geological Society of America Special Paper», 247, in stampa.

Un'eruzione vulcanica

Per spiegare la scomparsa di molte forme di vita avvenuta 65 milioni di anni fa numerosi indizi chiamano in causa un'intensa attività vulcanica

di Vincent E. Courtillot

La misteriosa estinzione in massa avvenuta 65 milioni di anni fa è stata attribuita ora all'impatto di un grosso asteroide, ora a un'imponente eruzione vulcanica. Entrambe le ipotesi presuppongono che nubi di polvere e cambiamenti chimici nell'atmosfera e negli oceani abbiano innescato una dilagante crisi ecologica che avrebbe cancellato un gran numero di famiglie di animali e di piante. La documentazione fossile, in generale, è compatibile con entrambe le ipotesi; la controversia principale riguarda la rapidità con cui si sarebbe consumato l'evento. Dati di nuova acquisizione fanno ritenere che l'estinzione in massa sia avvenuta in decine o persino in centinaia di migliaia di anni. Una tale durata corrisponde bene a un periodo di violente eruzioni vulcaniche avvenute in India appunto all'epoca dell'estinzione in massa. Inoltre, sembra che nella storia della Terra anche altre estinzioni siano coincise con periodi di intensa attività vulcanica.

Le divisioni convenzionali della storia della Terra riflettono forti discontinuità geologiche e biologiche. L'estinzione in massa di 65 milioni di anni fa segnò la fine dell'Era mesozoica, caratterizzata dal grande successo evolutivo dei rettili, e l'inizio dell'Era cenozoica, che avrebbe visto predominare di gran lunga i mammiferi. Dal momento che l'ultimo periodo del Mesozoico è il Cretaceo e il primo periodo del Cenozoico è il Terziario, l'epoca della più recente estinzione in massa viene definita limite Cretaceo--Terziario, o semplicemente limite KT.

In corrispondenza di questo limite i dinosauri trovarono la loro fine e, fatto ancora più straordinario, scomparve il 90 per cento di tutti i generi di protozoi e di alghe. John J. Sepkoski, Jr., e David M. Raup della Università di Chicago ritengono che a quell'epoca fosse scomparso dal 60 al 75 per cento di tutte le specie. Fatto di uguale importanza, sopravvissero molte specie, tra le quali i progenitori degli esseri umani.

Come è ampiamente descritto nell'articolo che precede il mio in questo dibattito, nel 1980 Luis W. e Walter Alvarez (padre e figlio) dell'Università della California a Berkeley, insieme con i colleghi Frank Asaro e Helen V. Michel, scoprirono in Italia, Danimarca e Nuova Zelanda, in rocce risalenti al limite KT, concentrazioni insolitamente alte di iridio, da 10 a 100 volte i normali livelli. L'iridio è raro nella crosta terrestre, ma può essere relativamente abbondante in altre parti del sistema solare. Perciò il gruppo di Berkeley concluse che l'iridio dovesse provenire dallo spazio esterno e formulò l'ipotesi della caduta di un enorme asteroide sulla Terra. Il colossale impatto avrebbe avvolto la Terra in una nube di polvere, provocando oscurità, interruzione della fotosintesi, collasso delle catene alimentari e, da ultimo, estinzioni in massa.

L'iridio anomalo è contenuto in un sottile strato d'argilla la cui composizione chimica differisce da quella degli strati al di sopra e al di sotto del limite KT. Il gruppo di Alvarez interpretò questa argilla come resto alterato della polvere sollevata dall'impatto. In quest'ottica lo strato del limite si sarebbe depositato in meno di un anno: meno di un batter d'occhio se paragonato ai tempi geologici. Anche altri ritrovamenti insoliti nello strato corrispondente al limite KT, in particolare cristalli di quarzo che sembra siano stati soggetti a sollecitazioni fisiche estremamente intense, si potrebbero spiegare con l'impatto di un asteroide.

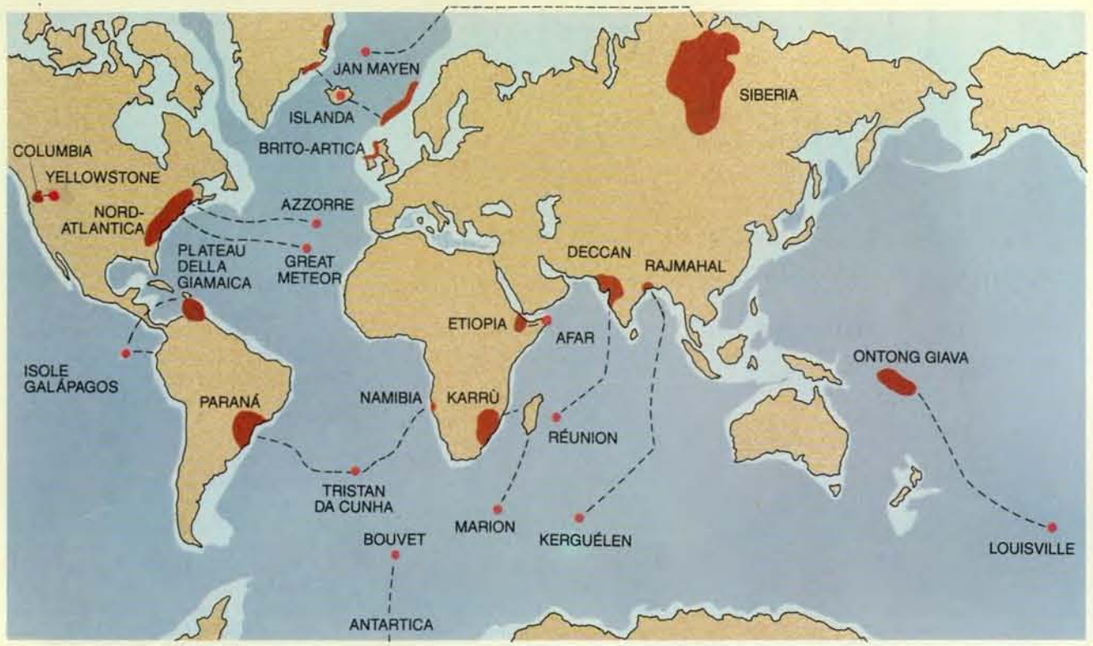
Da qualche tempo andava prendendo forma un'ipotesi alternativa a quella dell'asteroide. Già nel 1972 Peter R. Vogt del Naval Research Laboratory di Washington aveva fatto notare che un vulcanismo esteso si verificò, principalmente in India, circa nello stesso periodo in cui si era depositato lo strato corrispondente al limite KT. Il vulcanismo aveva prodotto vaste colate di lava, i trappi del Deccan (Deccan Traps). Vogt pensò che i trappi potessero essere collegati ai numerosi cambiamenti avvenuti alla fine del Cretaceo.

A metà degli anni settanta Dewey M. McLean del Virginia Polytechnic Institute ipotizzò che i vulcani avessero potuto causare le estinzioni in massa immettendo nell'atmosfera grandi quantità di anidride carbonica, il che avrebbe provocato repentini cambiamenti climatici e alterato la composizione chimica degli oceani. Charles B. Officer e Charles L. Drake del Dartmouth College analizzarono sedimenti provenienti da sezioni del limite KT concludendo che l'arricchimento in iridio e altre anomalie chimiche del limite non potevano essersi verificati istantaneamente, ma avevano richiesto un periodo di 10 000-100 000 anni. Secondo questi ricercatori le anomalie potevano più facilmente essere attribuite all'origine vulcanica che a quella meteoritica.

Il lasso di tempo rappresentato dallo strato d'argilla del limite KT ha sempre rappresentato uno degli aspetti più controversi. Non è ancora possibile datare una roccia di 100 milioni di anni con la precisione di una parte su 1000 (cioè fino a 100 000 anni), tuttavia gran parte del dibattito si concentra su una questione temporale: è possibile che l'argilla dello strato limite KT si sia depositata in meno di un anno (come ci si aspetterebbe da un impatto) oppure furono necessari almeno 10 000 anni (un lungo periodo di vulcanismo)?.

La sola estensione dei trappi del Deccan è sufficiente a far pensare che questa formazione deve essere stata di grande importanza nella storia della Terra: le singole colate di lava si estendono su una superficie di ben 10 000 chilometri quadrati e hanno un volume che supera i 10 000 chilometri cubi. Lo spessore delle colate è mediamente di 10-50 metri e qualche volta raggiunge i 150 metri. Nell'India occidentale la coltre di colate laviche raggiunge spesso i 2400 metri. Le colate possono aver inizialmente coperto più di due milioni di chilometri quadrati e il volume totale può aver superato due milioni di chilometri cubi.

Restava da chiarire se la durata del vulcanismo del Deccan fosse compatibile con l'età e lo spessore del limite KT. Fino a poco tempo fa si pensava che i campioni di lava dei trappi del Deccan avessero un'età variabile da 30 a 80 milioni di anni (valore stimato misurando nelle rocce il decadimento dell'isotopo radioattivo potassio 40). Non era noto se questo intervallo fosse reale o piuttosto riflettesse un errore di misurazione. Affascinato da questo problema, nel 1985 mi sono unito alla ricerca dei colleghi per trovare una risposta attendibile.



Le province basaltiche (o plateau basaltici) sono effusioni laviche che testimoniano passati sollevamenti geologici. Ogni provincia è correlata a un punto caldo, un sito di continua risalita magmatica che rimane fisso rispetto alle zolle litosferiche in movimento.

ESTINZIONE IN MASSA (Età approssimata in milioni di anni)	EFFUSIONE BASALTICA	PUNTO CALDO CORRELATO
MIOCENE MEDIO (14 ± 3)	COLUMBIA (16 ± 1)	YELLOWSTONE
MIOCENE INFERIORE (36 ± 2)	ETIOPIA (35 ± 2)	AFAR
MAASTRICHTIANO (65 ± 1) LIMITE KT	BRITO-ARTICA (62 ± 2)	ISLANDA
	TRAPPI DEL DECCAN (66 ± 2)	RÉUNION
CENOMANIANO (91 ± 2)	VULCANI DEL PACIFICO OCCIDENTALE (92 ± 3)	GRANDE ONDULAZIONE DEL PACIFICO
TITONIANO (137 ± 7)	PARANÁ (130 ± 5)	TRISTAN DA CUNHA
	NAMIBIA (135 ± 5)	
PLIENSBACHIANO (191 ± 3)	KARRÙ (190 ± 5)	MARION
NORICO (211 ± 8)	NORD ATLANTICA (200 ± 5)	AZZORRE/GREAT METEOR
DIULFIANO (249 ± 4) LIMITE PT	SIBERIA (250 ± 10)	JAN MAYEN

Un importante indizio è emerso dal fatto che le rocce del Deccan sono basalti, cioè rocce vulcaniche ricche di magnesio, titanio e ferro, discretamente magnetiche. Quando la lava basaltica si raffredda, nella roccia la magnetizzazione dei piccoli cristalli degli ossidi di ferro-titanio rimane «congelata», in posizione di allineamento con il campo magnetico terrestre. La polarità del campo si inverte casualmente cosicché il polo nord magnetico diventa polo sud e viceversa. Queste veloci inversioni - che durano circa 10 000 anni - avvengono in modo irregolare, a un ritmo che è variato da circa una sola inversione per milione di anni alla fine del Cretaceo ad approssimativamente quattro inversioni ogni milione di anni in tempi recenti.

Jean Besse e Didier Vandamme dell'Institut de Physique du Globe di Parigi e io abbiamo scoperto che più dell'80 per cento dei campioni di roccia provenienti dai trappi del Deccan presentava la stessa inversione di polarità. Se il vulcanismo fosse stato realmente ininterrotto tra 80 e 30 milioni di anni fa, si dovrebbe trovare un numero circa uguale di campioni a magnetizzazione normale e a magnetizzazione invertita, perché durante quel periodo di 50 milioni di anni avvennero decine di inversioni.

In realtà, le sezioni affioranti più potenti (1000 metri di spessore) dei trappi registrano solo una o due inversioni. Perciò nel 1986 abbiamo stabilito che che il vulcanismo del Deccan dovrebbe aver avuto inizio durante un intervallo di attività magnetica normale, avrebbe raggiunto il culmine nel successivo intervallo di polarità invertita e poi si sarebbe placato in un intervallo di polarità normale. Sulla base della frequenza usuale delle inversioni, i nostri risultati indicano che il vulcanismo non dovrebbe essere durato più di un milione di anni.

Se le cose stanno così, le età stabilite con la datazione mediante potassio 40 dovevano essere sbagliate. I miei colleghi Henri Maluski dell'Università di Montpellier e Gilbert Féraud dell'Università di Nizza e altri ricercatori hanno usato una tecnica più nuova e affidabile, la datazione argo-argo, per determinare quanto potassio 40 fosse decaduto nei campioni di roccia dai tempi della formazione. I loro risultati hanno confermato che le effusioni del Deccan si formarono in un periodo relativamente breve. Le stime delle età delle lave del Deccan vanno dai 64 ai 68 milioni di anni; variazioni in più o in meno possono dipendere da alterazione dei campioni o da differenti standard di laboratorio.

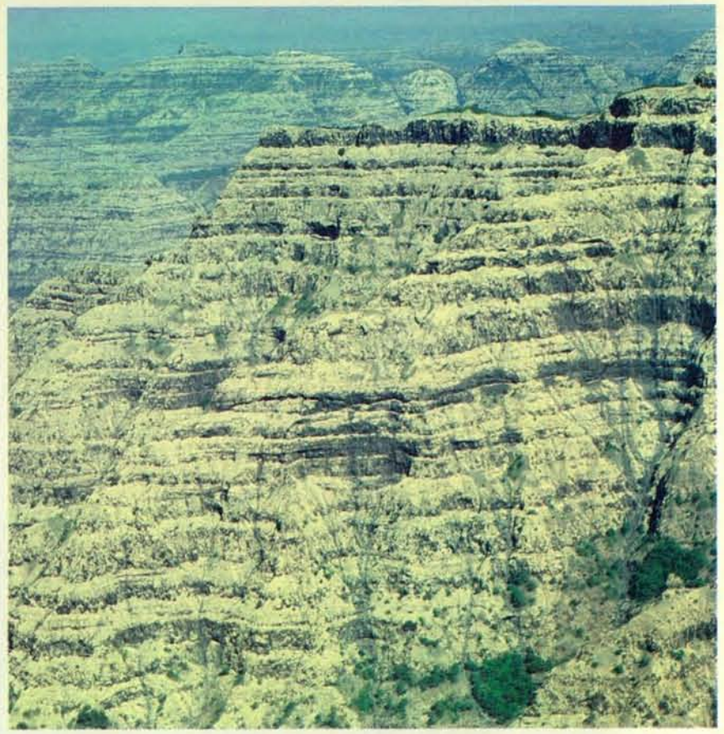
Sebbene sia difficile un'accurata datazione della roccia sedimentaria, le recenti scoperte effettuate da Ashok Sahni dell'Università di Chandigarh, J. J. Jaeger dell'Università di Montpellier e colleghi restringono ulteriormente le stime dell'età dei trappi del Deccan. Sedimenti immediatamente al di sotto delle colate del Deccan contengono frammenti di fossili di dinosauri che sembrano risalire al Maastrichtiano (ultimi otto milioni di anni del Cretaceo). Denti di dinosauri e di mammiferi e frammenti di uova di dinosauro che sembrano appartenere al Maastrichtiano sono stati trovati anche in strati sedimentari intercalati alle colate. Questo implica che il vulcanismo del Deccan abbia avuto inizio durante l'ultima fase del Cretaceo.

Dati più precisi provengono dai pozzi di esplorazione petrolifera che sono stati perforati lungo la costa orientale dell'India, dove hanno attraversato tre sottili colate basaltiche intercalate a strati di roccia sedimentaria. Il livello più basso di lava appoggia su strati sedimentari che contengono fossili della specie planctonica *Abatomphalus mayaroensis*, molto diffusa durante l'ultimo milione di anni del Cretaceo ed estintasi subito dopo. Gli strati di roccia sedimentaria intercalati alle effusioni laviche contengono anche fossili esattamente della stessa epoca, ma gli strati al di sopra delle effusioni sono sterili.

I fossili di *A. mayaroensis* compaiono in strati con polarità magnetica normale che giacciono al di sotto (e sono quindi più antichi) del limite KT e scompaiono in coincidenza con il limite stesso, che è localizzato nel successivo insieme di strati a polarità inversa (*si veda l'illustrazione a pagina 52*).

La conclusione più ragionevole che si può trarre dalle diverse prove è che il vulcanismo del Deccan debba essere incominciato durante l'ultimo intervallo magnetico normale del Cretaceo, abbia avuto il suo culmine durante il successivo intervallo a magnetizzazione inversa (al limite Cretaceo-Terziario o molto vicino) e sia terminato nel primo intervallo magnetico normale dell'Era cenozoica.

Gli studi magnetici e paleontologici concordano nel circoscrivere la durata del vulcanismo del Deccan a circa 500 000 anni, la migliore risoluzione di tempo ottenibile usando le tecniche attuali. Il fatto che il vulcanismo del Deccan - uno degli episodi più estesi e più rapidi di effusione lavica degli ultimi 250 milioni di anni - sia coinciso con il limite KT (nell'ambito delle più precise cronologie oggi raggiungibili) ci ha reso difficile evitare di concludere che esista un legame tra il vulcanismo dei trappi del Deccan e le estinzioni in massa.



I trappi del Deccan, in India, sono effusioni basaltiche stratificate databili al limite Cretaceo-Terziario, in concomitanza con l'ultima estinzione in massa della storia geologica.

Avendo stabilito che i trappi del Deccan vennero effusi quasi simultaneamente all'estinzione della fine del Cretaceo, abbiamo cercato in seguito di appurare se una eruzione vulcanica potesse spiegare le strutture osservate negli strati del limite KT. In effetti queste strutture potrebbero essere state verosimilmente prodotte sia da un'immane eruzione vulcanica, sia dall'impatto di un asteroide.

Non è detto che l'inconsueto strato ricco d'iridio che sembra essersi depositato simultaneamente su tutta la Terra debba necessariamente provenire dallo spazio. William H. Zoller, Ilhan Olmez e colleghi dell'Università del Maryland a College Park hanno scoperto anomali aumenti di iridio in particelle emesse dal vulcano Kilauea, nell'isola di Hawaii. J. P. Toutain e G. Meyer dell'Institut de Physique du Globe hanno trovato iridio in particelle emesse da un altro vulcano, il Piton de la Fournaise, sull'isola di Réunion che, come diremo più avanti, è correlato al vulcanismo del Deccan. Polvere vulcanica ricca di iridio è stata trovata inclusa nella calotta glaciale antartica, a migliaia di chilometri dai vulcani che l'avevano emessa.

La composizione mineralogica dell'argilla dello strato limite differisce da quella delle argille presenti al di sopra e al di sotto dello strato. L'illite, tipico minerale argilloso, è sostituita dalla smectite che si può formare per alterazione della roccia basaltica. Recenti studi mineralogici sull'argilla del limite KT a Stevns Klint in Danimarca hanno portato W.

Crawford Elliott della Case Western Reserve University e collaboratori e Birger Schmitz dell'Università di Göteborg a concludere che l'argilla è costituita da un tipo particolare di smectite che in realtà è cenere vulcanica alterata.

L'argilla del limite KT può essere simulata mescolando 10 parti di materiale della crosta terrestre con una parte di materiale delle comuni meteoriti litoidi. Tuttavia il mantello terrestre ha una composizione simile a quella delle meteoriti, cosicché potrebbe dar origine a uguali anomalie chimiche. Karl K. Turekian della Yale University e Jean-Marc Luck dell'Institut de Physique du Globe hanno trovato che l'abbondanza relativa degli elementi renio e osmio nell'argilla è simile a quella delle meteoriti e del mantello terrestre.

Le peculiari strutture fisiche del materiale proveniente dal limite KT possono essere anche spiegate da entrambe le ipotesi. Gli strati del limite contengono una gran quantità di minuscole sferule. Alcune di esse sono costituite da minerali argillosi che sembrano resti alterati di goccioline di basalto, ma è impossibile affermare se si siano originate come prodotti piroclastici oppure da crosta oceanica fusa in seguito all'impatto di un asteroide. Inoltre contribuisce a confondere le idee il fatto che almeno alcune delle sferule si sono rivelate alghe fossili tondeggianti o perfino uova di insetti recenti che hanno contaminato il materiale.

La scoperta negli strati del limite KT di granuli di quarzo deformati con struttura da shock, fatta per la prima volta da Bruce F. Bohor e Glen A. Izett dello US Geological Survey di Denver, è spesso considerata come la prova più convincente a favore dell'ipotesi dell'impatto. Questi granuli deformati sono stati trovati in precedenza solo nei crateri da impatto noti (come il Meteor Crater dell'Arizona) o in luoghi di esplosioni nucleari sotterranee. Essi si sono formati per intensa sollecitazione dinamica a pressioni 100 000 volte superiori a quella atmosferica, ma strutture da shock si possono formare anche a pressioni molto più basse se la roccia si riscalda prima che avvenga la sollecitazione, come nel caso di un'eruzione vulcanica.

Quando il magma risale alla superficie terrestre si decomprime e libera i gas. Contemporaneamente esso si raffredda e solidifica, ma se il raffreddamento è troppo rapido i gas rimangono intrappolati nel magma e la pressione aumenta finendo col provocare esplosioni e forti onde d'urto. Queste sollecitazioni potrebbero bastare per formare cristalli di quarzo deformati se la temperatura e la durata del fenomeno fossero sufficienti.

Un magma ricco di silicati è viscoso e particolarmente adatto a provocare eruzioni esplosive; un vulcanismo di questo tipo caratterizza il Vesuvio e il St. Helens. Nel 1986 Neville L. Carter della Texas A&M University e collaboratori hanno trovato prove di strutture da shock simili a quelle del limite KT in rocce provenienti da località che sono state sede di vulcanismo esplosivo recente, come la grande eruzione di 75 000 anni fa a Toba, nell'isola di Sumatra. Usando il microscopio elettronico a trasmissione Jean-Claude Doukhan dell'Università di Lille ha recentemente scoperto che le strutture da shock prodotte in laboratorio, quelle da impatto meteoritico e quelle osservate in campioni provenienti dal limite KT differiscono tutte tra di loro per alcuni aspetti; inoltre pare che la somiglianza tra le strutture ottenute in laboratorio e quelle da impatto sia stata sopravvalutata. Le strutture da shock dei campioni KT sono caratterizzate da bollicine microscopiche che non si osservano nei campioni collegabili a impatti e che sembrano indicare una temperatura di formazione più elevata, compatibile con un'origine vulcanica.

Il vulcanismo silicico esplosivo di solito precede periodi di relativa quiete, come quello che contrassegna il vulcanismo «tipo Deccan» (a colate basaltiche) durante il quale la lava basaltica fluisce liberamente e abbondantemente. Il 10-15 per cento del volume di lava delle colate di tipo Deccan fuoriesce con episodi di vulcanismo silicico esplosivo dovuto al fatto che un pennacchio ascendente di magma caldo si apre la strada attraverso la crosta continentale dando origine a magmi viscosi ricchi di silice (acidi).

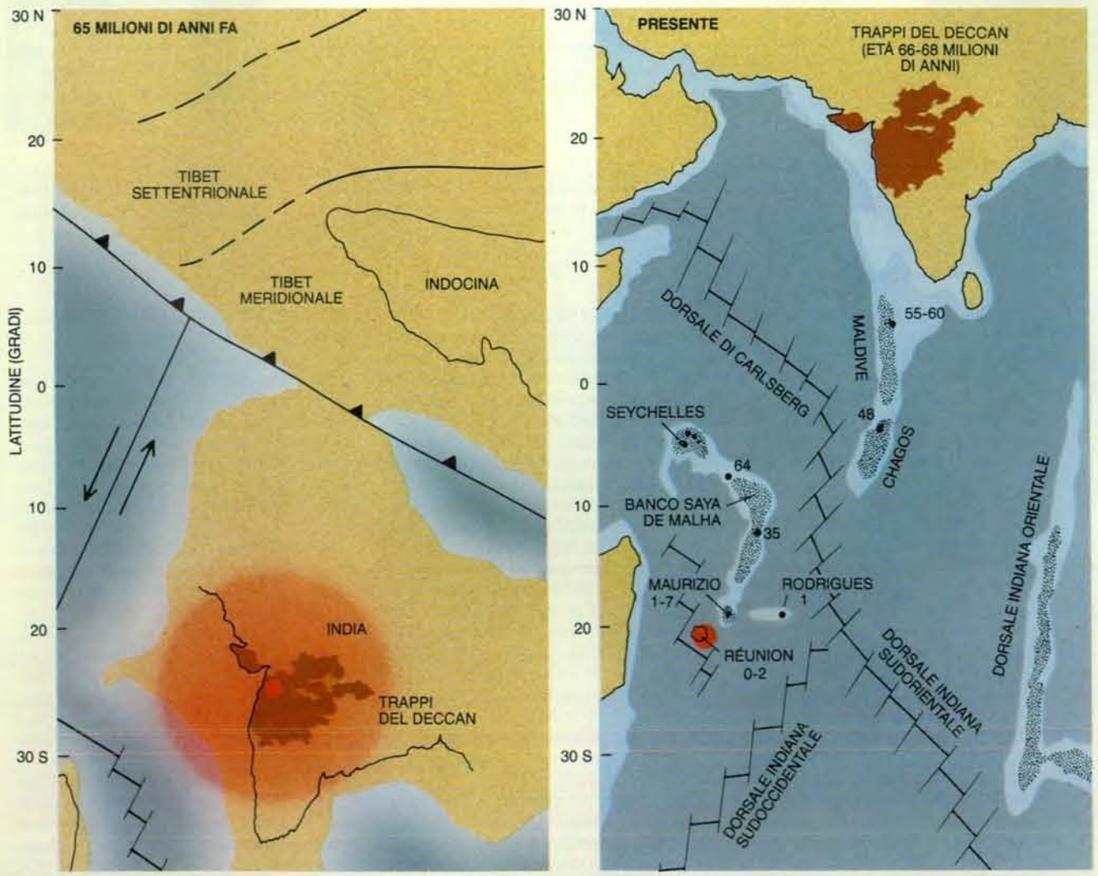
Le inusuali strutture chimiche e fisiche negli strati del limite KT sono presenti in tutto il mondo. Un impatto di asteroide potrebbe aver lanciato materiale nella stratosfera, da dove esso sarebbe ricaduto in ogni parte del globo. D'altra parte, Richard B. Stothers e collaboratori del Goddard Space Flight Center della NASA di Greenbelt, nel Maryland, hanno elaborato un modello che rappresenta il modo in cui le fontane di lava, come quelle del Kilauea nell'isola di Hawaii, eruttano polveri e prodotti piroclastici. Rapportando il modello alle dimensioni del vulcanismo del Deccan, questi autori hanno stabilito che un'effusione lavica estesa come quella dovrebbe avere immesso nella stratosfera una grande quantità di materiale. La circolazione atmosferica potrebbe poi aver distribuito il materiale piuttosto uniformemente sui due emisferi, indipendentemente dalla provenienza.

Le spaventose conseguenze dell'impatto di un asteroide e di un'immane attività vulcanica sarebbero molto simili. La prima sarebbe stata l'oscurità conseguente all'immissione di grandi quantità di polveri nell'atmosfera (sia come materiale sollevato dall'impatto sia come cenere vulcanica). L'oscurità avrebbe bloccato la fotosintesi interrompendo le catene alimentari. Questo disastro ambientale appare documentato dai resti fossili. Gli organismi di acqua dolce furono molto meno colpiti di quelli terrestri o marini forse perché gli animali di acqua dolce non si nutrivano di piante vascolari (come fanno molti animali terricoli) o di plancton fotosintetico (una importante fonte alimentare per i vertebrati marini che fu distrutta alla fine del Cretaceo).

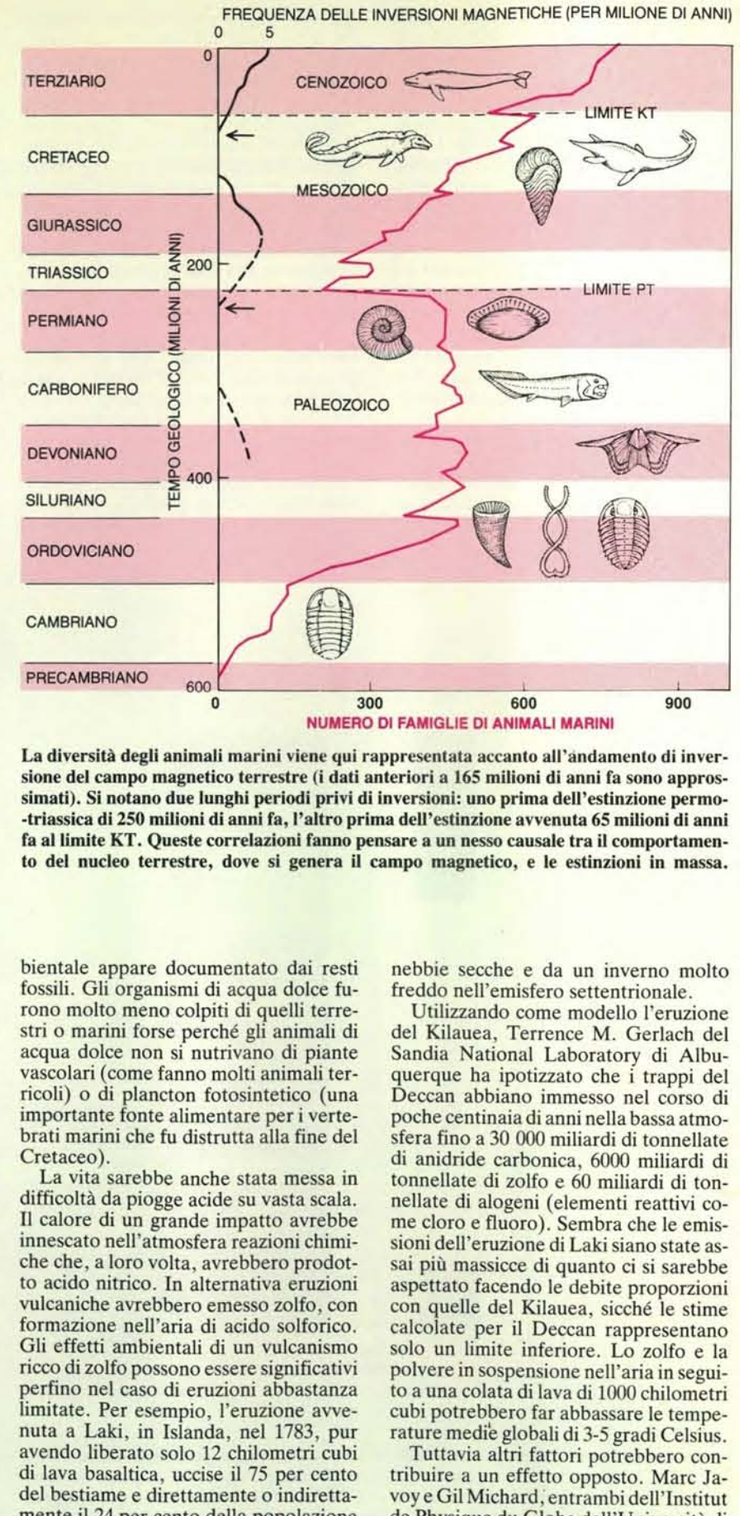
La vita sarebbe anche stata messa in difficoltà da piogge acide su vasta scala. Il calore di un grande impatto avrebbe innescato nell'atmosfera reazioni chimiche che, a loro volta, avrebbero prodotto acido nitrico. In alternativa eruzioni vulcaniche avrebbero emesso zolfo, con formazione nell'aria di acido solforico. Gli effetti ambientali di un vulcanismo ricco di zolfo possono essere significativi perfino nel caso di eruzioni abbastanza limitate. Per esempio, l'eruzione avvenuta a Laki, in Islanda, nel 1783, pur avendo liberato solo 12 chilometri cubi di lava basaltica, uccise il 75 per cento del bestiame e direttamente o indirettamente il 24 per cento della popolazione del paese. L'evento fu seguito da strane nebbie secche e da un inverno molto freddo nell'emisfero settentrionale.

Utilizzando come modello l'eruzione del Kilauea, Terrence M. Gerlach del Sandia National Laboratory di Albuquerque ha ipotizzato che i trappi del Deccan abbiano immesso nel corso di poche centinaia di anni nella bassa atmosfera fino a 30 000 miliardi di tonnellate di anidride carbonica, 6000 miliardi di tonnellate di zolfo e 60 miliardi di tonnellate di alogeni (elementi reattivi come cloro e fluoro). Sembra che le emissioni dell'eruzione di Laki siano state assai più massicce di quanto ci si sarebbe aspettato facendo le debite proporzioni con quelle del Kilauea, sicché le stime calcolate per il Deccan rappresentano solo un limite inferiore. Lo zolfo e la polvere in sospensione nell'aria in seguito a una colata di lava di 1000 chilometri cubi potrebbero far abbassare le temperature medie globali di 3-5 gradi Celsius.

Tuttavia altri fattori potrebbero contribuire a un effetto opposto. Marc Javoy e Gil Michard, entrambi dell'Institut de Physique du Globe dell'Università di Parigi ipotizzano che l'anidride solforosa



Secondo l'autore (*schema a sinistra*), il punto caldo attualmente individuato sotto il vulcano dell'isola di Réunion (il Piton de la Fournaise) si trovava sotto l'India quando i trappi del Deccan vennero effusi. Le età dei vulcani sottomarini a sud dell'India aumentano regolarmente da Réunion ai trappi del Deccan, dove la linea dell'attività vulcanica ha fatto la sua comparsa (*a destra*).



La diversità degli animali marini viene qui rappresentata accanto all'andamento di inversione del campo magnetico terrestre (i dati anteriori a 165 milioni di anni fa sono approssimati). Si notano due lunghi periodi privi di inversioni: uno prima dell'estinzione permo-triassica di 250 milioni di anni fa, l'altro prima dell'estinzione avvenuta 65 milioni di anni fa al limite KT. Queste correlazioni fanno pensare a un nesso causale tra il comportamento del nucleo terrestre, dove si genera il campo magnetico, e le estinzioni in massa.

dei vulcani del Deccan abbia reso acida la superficie dell'oceano, distruggendo le alghe che, in condizioni normali, assumono anidride carbonica dall'atmosfera e poi, alla morte, la trasferiscono sul fondo dell'oceano. Anche le acque acide dell'oceano avrebbero disciolto i sedimenti carbonatici del fondo, liberando l'anidride carbonica in essi intrappolata. Nel complesso i livelli atmosferici di anidride carbonica sarebbero cresciuti rapidamente fino a circa otto volte la concentrazione attuale provocando un innalzamento della temperatura di cinque gradi Celsius. Non è chiara l'interazione tra il raffreddamento dovuto alle polveri e il riscaldamento da anidride carbonica (fenomeni che possono avvenire su scale di tempo ampiamente differenti), ma i cambiamenti climatici risultanti devono probabilmente essere stati particolarmente traumatici per l'ecosistema globale. Entrambe le ipotesi - asteroide o vulcanismo - implicano analoghi effetti di raffreddamento e di successivo riscaldamento.

Finora le prove invocate nel dibattito sono state ugualmente compatibili con entrambe le ipotesi. Ma molti dettagli fanno pensare che le estinzioni in massa e gli anomali processi fisici che si verificarono alla fine del Cretaceo si siano svolti nel corso di centinaia di migliaia di anni. Questo periodo è paragonabile alla durata del vulcanismo del Deccan, ma è incompatibile con un improvviso impatto da parte di un asteroide.

Alcuni paleontologi sostengono che l'estinzione alla fine del Cretaceo non sia stata un singolo istantaneo evento. Sembra che i tassi d'estinzione abbiano cominciato ad aumentare già un milione di anni prima del limite KT. Persino in prossimità del limite il modello non è uniforme: per esempio, alcune specie di foraminiferi planctonici e di nanoplancton (alghe calcaree microscopiche) mostrano differenti andamenti di estinzione e di ripresa. Questo fenomeno è definito estinzione in massa graduale.

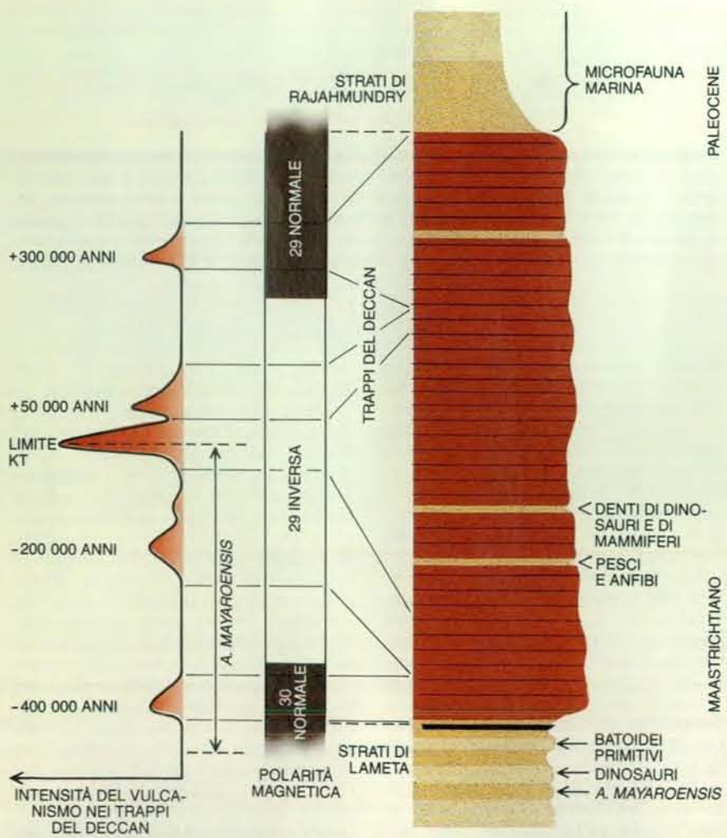
Uno degli studi recenti più accurati sulla modalità con cui sarebbero avvenute le estinzioni è stato condotto da Gerta Keller della Princeton University. Analizzando le sezioni ben conservate del limite KT in Tunisia e nel Texas, la Keller trovò la prova di una prima fase di estinzione (vista anche nei macrofossili) che cominciò 300 000 anni prima dell'evento dell'iridio KT e di un'altra estinzione che ebbe luogo 50 000 anni dopo il limite. La Keller attribuisce il primo evento all'abbassamento del livello del mare e al globale raffreddamento.

Un altro fatto conferma che la Terra debba aver subito alla fine del Cretaceo non uno, ma molti sconvolgimenti. Sappiamo, per esempio, che si verificò un cambiamento improvviso nella quantità di carbonio 13 e di ossigeno 18 (isotopi rispettivamente leggero e pesante di questi elementi, le cui concentrazioni variano in relazione alla temperatura e all'acidità dell'oceano e al numero di organismi in esso presenti). Le estinzioni e le oscillazioni del carbonio 13 osservate in strati della Spagna si trovano in intervalli di rocce magnetiche che corrispondono allo stesso modello di polarità normale-invertita-normale riscontrato nei trappi del Deccan.

Perfino l'iridio sembra mostrare un certo numero di piccole oscillazioni in prossimità del limite KT. Robert Rocchia e colleghi del Commissariat à l'énergie atomique e del CNRS di Gif-sur-Yvette e di Saclay, in Francia, hanno trovato picchi secondari di iridio al di sopra e al di sotto dello strato primario di iridio (corrispondenti a intervalli di tempo di circa 10 000 anni) nell'argilla del limite KT in Spagna e in Danimarca. Rocchia, i suoi colleghi e io abbiamo trovato che lo strato ricco di iridio a Gubbio, in Italia, sembra estendersi per circa 500 000 anni. Gli assai discussi cristalli di quarzo deformati mostrano una distribuzione analoga. Officer e Carter hanno scoperto che questi cristalli si estendono per quattro metri nella sezione di Gubbio, il che corrisponde ancora a un arco di tempo di circa 500 000 anni.

James C. Zachos dell'Università del Rhode Island e collaboratori hanno rilevato la composizione chimica di microfossili provenienti dal fondo del Pacifico settentrionale e hanno trovato che la produttività biologica in mare aperto fu interrotta al tempo del limite KT e per circa altri 500 000 anni. Hanno anche dedotto che alcuni cambiamenti ambientali significativi, compreso il raffreddamento, devono avere avuto inizio almeno 200 000 anni prima del limite stesso.

Alcuni sostenitori della teoria dell'impatto, in particolare Piet Hut dell'Institute for Advanced Study di Princeton e colleghi, per spiegare queste scoperte hanno prontamente sostituito con una serie di impatti di comete l'impatto di un solo asteroide. La ricerca di una risposta globale ha condotto anche all'ipotesi che i trappi del Deccan segnino il luogo dell'impatto dell'asteroide, ma sembra difficile dimostrarlo in quanto in India non è stata trovata alcuna traccia dell'ipotetico impatto. Robert S. White dell'Università di Cambridge ha inoltre dimostrato che impatti di grande entità non sono in grado di innescare un imponente vulcanismo, dal momento che la parte del mantello appena sotto la litosfera non contiene normalmente grandi serbatoi di roccia fusa. Per di più il vulcanismo del Deccan ebbe inizio durante un intervallo geomagnetico normale, poche centinaia di migliaia di anni prima dell'intervallo magnetico invertito che racchiude l'anomalia dell'iridio KT e lo strato d'argilla.



Le rocce sedimentarie intercalate ai basalti del Deccan contengono resti di dinosauri e altri fossili del Cretaceo, assenti negli strati sovrastanti i basalti. L'estinzione in massa avvenne mentre le eruzioni erano in corso (*a destra*). La storia vulcanica (*a sinistra*) può spiegare le estinzioni multiple e le anomalie chimiche delle rocce della fine del Cretaceo. Le inversioni magnetiche contribuiscono a sincronizzare la stratigrafia con il vulcanismo ipotizzato.

Durante il Cretaceo il vulcanismo aumentò di intensità, il livello del mare si alzò e poi si abbassò drasticamente e tutto il mantello subì un significativo spostamento. Nel Cretaceo e nel periodo che lo precedette, il Giurassico, avvennero anche i maggiori smembramenti dei continenti. Tra 120 e 85 milioni di anni fa il campo magnetico terrestre non subì inversioni, ma 15-20 milioni di anni prima del limite KT il campo incominciò nuovamente a invertirsi. La frequenza delle inversioni, che indica attività nel nucleo terrestre e al limite mantello-nucleo, è aumentata regolarmente da allora a oggi, fino a circa una volta ogni 250 000 anni.

Tutti questi aspetti possono essere messi in relazione con un episodio di attiva convezione del mantello che cominciò decine di milioni di anni prima del limite KT. Secondo me, l'esistenza di anomalie geodinamiche, geologiche e paleontologiche sovrapposte a breve e lungo termine indica una causa interna comune.

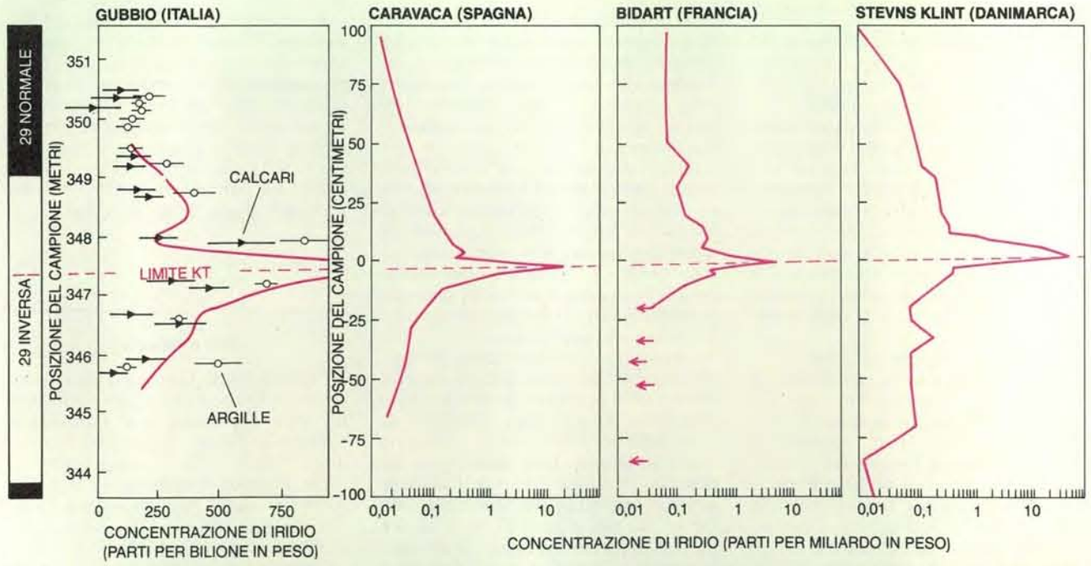
Quale potrebbe essere questa causa? Una risposta verosimile viene dalla teoria dei punti caldi del mantello sviluppata da W. Jason Morgan della Princeton University e da altri. Peter L. Olson e Harvey Singer della Johns Hopkins University hanno sviluppato un modello che può spiegare queste zone di attività vulcanica persistente. Nel modello un pennacchio di materiale caldo, a bassa densità e viscosità, sale dalle parti più profonde del mantello, formando una testa quasi sferica mentre si fa strada attraverso la regione del mantello più fredda e densa. La testa continua a crescere fintantoché la roccia fusa ascendente la alimenta.

White e Dan P. McKenzie, anch'egli di Cambridge, Mark Richards e Robert A. Duncan dell'Oregon State University ritengono, come l'autore dell'articolo, che, quando un pennacchio si innalza dal mantello, la crosta al di sopra di esso si inarchi e si distenda dando origine alla formazione di rift continentali (si veda *Vulcanismo di rift* dei primi due autori qui citati, in «Le Scienze» n. 253, settembre 1989). Secondo questa ipotesi il materiale del pennacchio giunto in prossimità della superficie si decomprimerebbe, fondendo così rapidamente (in tempi geologici: meno di un milione di anni). La sommità del pennacchio solleverebbe quindi un'estesa area crostale cosicché, se il magma riuscisse a raggiungere la superficie, potrebbe scorrere lungo i pendii, dando luogo a estese effusioni.

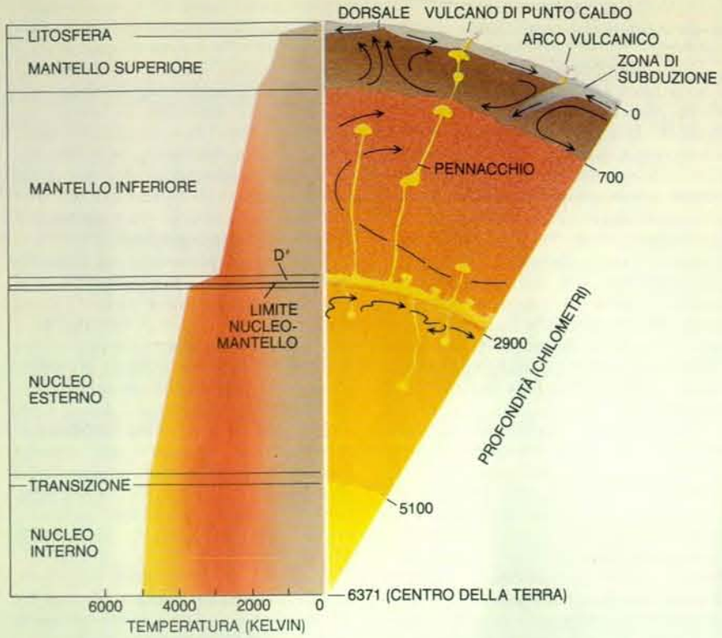
I trappi del Deccan sarebbero il risultato dell'arrivo di un pennacchio alla base della litosfera. Il vulcanismo provocato da un pennacchio caldo sarebbe rapido e altamente episodico. Le singole colate verrebbero effuse in giorni o settimane; la colata successiva potrebbe avvenire anche migliaia di anni più tardi. Le conseguenze ecologiche di vasta portata di ciascuna colata spiegherebbero le estinzioni in massa graduali.

Il gigantesco pennacchio di mantello che ha prodotto i trappi del Deccan avrebbe lasciato relitti strutturali e dinamici. Nel 1987 l'Ocean Drilling Program, diretto da Duncan, ha esplorato e datato una catena sottomarina di vulcani che si estende dall'India sudoccidentale, vicino ai trappi del Deccan, fino a Réunion, il vulcano attivo a est del Madagascar. Réunion è un punto caldo, attivato da un profondo flusso di magma caldo proveniente dal mantello, che si è fatto strada attraverso i continenti indiano e africano quando questi, durante la loro deriva, vi passarono sopra. L'età dei vulcani sottomarini di Réunion passa progressivamente da 0-2 milioni di anni intorno a Réunion fino a 55-60 milioni di anni a sud dei trappi del Deccan.

Richards, Duncan e io riteniamo che il punto caldo di Réunion rappresenti la coda di magma caldo che dovrebbe seguire il pennacchio responsabile dei trappi. Besse, Vandamme e io abbiamo verificato che il punto caldo attualmente sotto Réunion era localizzato alla fine del Cretaceo sotto i trappi del Deccan.



L'iridio rinvenuto nelle argille delle località qui considerate potrebbe essere stato liberato o dall'impatto di un asteroide o da un'enorme eruzione vulcanica. L'elevata concentrazione di iridio negli strati del limite KT in Italia, Spagna, Francia e Danimarca copre un periodo di 500 000 anni, mentre l'impatto di un asteroide avrebbe prodotto un improvviso aumento di iridio ben circoscritto nel tempo. Invece pare che le eruzioni del Deccan siano durate proprio il tempo giusto per spiegare la distribuzione dell'iridio.



Il modello del pennacchio del mantello è illustrato in questa sezione schematica della Terra. Il calore proveniente dal nucleo esterno ispessisce lo strato più basso del mantello, lo strato D″, annullando il moto convettivo del nucleo che determina le inversioni del campo magnetico. Alla fine lo strato D″ «esplode» in enormi pennacchi ascendenti. L'assenza di inversioni magnetiche durante il Cretaceo può indicare un accrescimento dello strato D″ che avrebbe poi dato origine ai pennacchi caldi i quali, raggiunta la superficie 65 milioni di anni fa, provocarono il catastrofico vulcanismo che formò i trappi del Deccan. Il punto caldo Piton de la Fournaise, sotto Réunion, potrebbe costituire la «coda» del pennacchio.

Non c'è traccia di punto caldo prima del limite KT; sembra che l'episodio di vulcanismo violento del Deccan segni l'arrivo del punto caldo alla superficie.

L'attività geologica interna associata alla risalita di un pennacchio di mantello si accorda con il comportamento del campo magnetico terrestre all'epoca del limite KT. Si pensa che la lenta convezione del ferro fuso nel nucleo terrestre esterno - 10 chilometri all'anno - dia origine al campo magnetico terrestre. Le instabilità al limite tra nucleo e mantello potrebbero essere la causa delle inversioni magnetiche.

Il calore proveniente dal nucleo innalza la temperatura (e quindi abbassa la densità) del materiale nello strato più profondo del mantello, lo strato D″, che aumenta di spessore fino a divenire instabile e a formare i pennacchi. Lunghi periodi di tempo con poche o nessuna inversione magnetica, come l'intervallo trascorso fra 120 e 85 milioni di anni fa, indicano assenza di attività del nucleo esterno e la crescita dello strato D″.

Circa 80 milioni di anni fa lo strato esplose in enormi pennacchi ascendenti di magma caldo. A questo punto, il flusso di calore dal nucleo al mantello sarebbe aumentato e le inversioni magnetiche sarebbero ricominciate. I pennacchi, alla velocità tipica che hanno nel mantello, circa un metro all'anno, avrebbero impiegato alcuni milioni di anni prima di raggiungere la superficie, dove la loro subitanea decompressione avrebbe provocato un vulcanismo esplosivo seguito da effusioni di lava. Pennacchi secondari più piccoli non avrebbero raggiunto la superficie, ma avrebbero accelerato la convezione del mantello, l'espansione dei fondi oceanici, le variazioni del livello del mare e gli altri eventi geologici che si verificarono nel Cretaceo.

Questo tipo di sollevamento geologico può essere una conseguenza naturale del fatto che la Terra è un'attiva e complessa macchina termica, costituita da strati dalle proprietà fisiche e chimiche molto differenti. Una convezione del mantello uniforme e ben regolata e la violenta instabilità dei pennacchi sono forse i due tipi estremi di modalità con cui il calore terrestre sfugge all'esterno. Se questo è davvero il «modo di funzionare» della Terra, catastrofi simili sarebbero già avvenute. In realtà i principali eventi di estinzione relativamente recenti (quelli che contrassegnarono la transizione tra Paleozoico e Mesozoico cominciarono 250 milioni di anni fa) sembrano correlati a epoche di estese effusioni basaltiche. Il fatto interessante è che il periodo più lungo conosciuto durante il quale non si ebbero inversioni del campo magnetico terrestre terminò anche con la più grande estinzione in massa, quella che segnò l'inizio dell'Era mesozoica. In quell'epoca più del 95 per cento delle specie marine scomparve. I trappi siberiani di 250 milioni di anni fa sono i primi chiamati in causa per questa estinzione.

Entrambe le ipotesi - impatto di un asteroide e vulcanismo - implicano che le catastrofi repentine possano avere grande importanza nell'evoluzione biologica. Questa opinione sembrerebbe in contraddizione con il concetto di attualismo, principio guida della geologia in base al quale la storia della Terra sarebbe stata determinata dai medesimi processi geologici che avvengono tuttora e agiscono su lunghi archi di tempo. Qualitativamente, eruzioni vulcaniche e impatti meteoritici non sarebbero eventi eccezionali, ma su base quantitativa, l'evento che portò all'estinzione dei dinosauri è certamente diverso da ogni altro, almeno negli ultimi 250 milioni di anni.

Le inversioni magnetiche del nucleo terrestre e l'ascesa di grandi pennacchi dal mantello possono essere manifestazioni del fatto che la Terra è un sistema caotico. Le variazioni nella frequenza delle inversioni magnetiche e lo smembramento dei continenti nelle ultime centinaia di milioni di anni mostrano una certa periodicità: sembra che gli episodi vulcanici catastrofici si verifichino a intervalli di 200 milioni di anni, con eventi minori circa ogni 30 milioni di anni.

Si sta ipotizzando che l'inizio del Paleozoico, 570 milioni di anni fa, quando comparve la vita pluricellulare, coincida con un episodio di questo tipo. In effetti le grandi estinzioni aprono improvvisamente ampi spazi ecologici che consentono ai nuovi organismi di svilupparsi. Gli eventi catastrofici potrebbero in realtà essere stati indispensabili per l'evoluzione di forme di vita superiori.

BIBLIOGRAFIA

OFFICER CHARLES B., HALLAM ANTHONY, DRAKE CHARLES L. e DEVINE JOSEPH D., *Late Cretaceous and Paroxysmal Cretaceous/Tertiary Extinctions* in «Nature», 326, n. 6109, 12 marzo 1987.

COURTILLOT V. e BESSE J., *Magnetic Field Reversals, Polar Wind and Core-Mantle Coupling* in «Science», 237, n. 4819, 4 settembre 1987.

COURTILLOT V., FÉRAUD G., MALUSKI H., VANDAMME D., MOREAU M. G. e BESSE J., *The Deccan Flood Basalts and the Cretaceous/Tertiary Boundary* in «Nature», 333, n. 6176, 30 giugno 1988.

Opere di consultazione

ENCICLOPEDIA DELLE ERBE
a cura di Richard Mabey *48 000 lire*

IL MANUALE DI GIARDINAGGIO
L'enciclopedia di tecniche di giardinaggio della Royal Horticultural Society
a cura di Christopher Brickell
48 000 lire

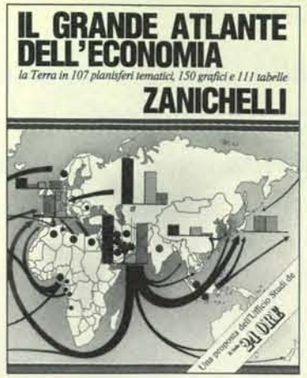
Giardinaggio

LE ROSE DEL XX SECOLO
di Peter Beales *78 000 lire*

Cucina

COLIN SPENCER
LA NUOVA CUCINA VEGETARIANA *48 000 lire*

Opere di consultazione



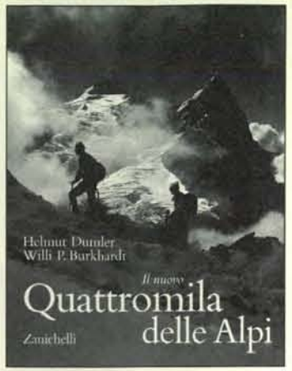
IL GRANDE ATLANTE DELL'ECONOMIA La Terra in 107 planisferi tematici, 150 grafici e 111 tabelle *48 000 lire*

IL GRANDE LIBRO DELLA STORIA ANTICA
di Daniele Manacorda e Giuseppe Pucci
74 000 lire

BIOLOGIA E MEDICINA
Biology & Medicine
Dizionario enciclopedico di scienze biologiche e mediche italiano-inglese inglese-italiano
di G. Delfino, E. Lanciotti
G. Liguri, M. Stefani
74 000 lire

LA CHIMICA DI TUTTI I GIORNI
di G. Vollmer,
M. Franz
34 000 lire
UN PRONTUARIO GUIDA PER IMPARARE A CONOSCERE E CONSUMARE I MILLE PRODOTTI DI USO QUOTIDIANO

Montagna



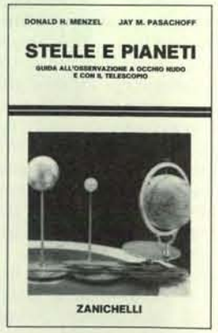
DUMLER, BURKHARDT
IL NUOVO QUATTROMILA DELLE ALPI
62 000 lire

CANZIANI, COLNAGO
GUIDA ALLA NATURA DEL GRAN PARADISO
36 000 lire

CARNISIO,
LAZZARIN, SOSTER
GUIDA ALLA VALSESIA
Arte e natura
36 000 lire

BERNETT,
GÖTZFRIED, ZINTL
PRONTO SOCCORSO IN MONTAGNA
Scuola di montagna 7, 25 000 lire

Guide



MENZEL, PASACHOFF
STELLE E PIANETI
Guida all'osservazione a occhio nudo e con il telescopio
58 000 lire

Zanichelli

Fotorifrazione laser

Nell'attraversare un cristallo, un raggio laser può dare origine a un improvviso sprazzo di luce: un fenomeno di fotorifrazione che potrebbe essere sfruttato per il funzionamento di calcolatori ottici superveloci

di David M. Pepper, Jack Feinberg e Nikolaj V. Kukhtarev

Quando circa 25 anni fa Arthur Ashkin e colleghi scoprirono presso i Bell Laboratories il fenomeno della fotorifrazione, lo considerarono una curiosità nel migliore dei casi e una grande iattura nel peggiore. Oggi i materiali fotorifrangenti cominciano a essere impiegati per costruire i componenti di una nuova generazione di calcolatori, in grado di sfruttare la luce anziché l'elettricità.

Ashkin effettuava i suoi esperimenti con un cristallo di niobato di litio ($LiNbO_3$), sperando di riuscire, con l'uso di questo composto, a trasformare il colore di un'intensa luce laser in un altro colore (processo che in gergo viene chiamato «generazione della seconda armonica»). Tra le altre prove, fece passare un fascio laser attraverso il cristallo; sulle prime questo si comportò benissimo, lasciando passare la luce del tutto indisturbata. Ma dopo pochi minuti il cristallo cominciò a distorcere il fascio, diffondendo luce per tutto il laboratorio. La luce laser doveva avere in qualche modo alterato le proprietà ottiche del cristallo. Questo effetto di fotorifrazione persisteva per alcuni giorni, ma se il cristallo veniva immerso in un fascio uniforme di luce riprendeva nuovamente a trasmettere fasci non distorti.

Negli ultimi 25 anni gli studiosi hanno scoperto un'ampia gamma di materiali fotorifrangenti, tra cui isolanti, semiconduttori e composti organici. Come le emulsioni delle pellicole fotografiche, i materiali fotorifrangenti si alterano rapidamente quando sono esposti a luce intensa, reagiscono lentamente quando sono sottoposti a luce fioca e riproducono minuti particolari quando sono colpiti da una figura luminosa complessa. A differenza di quanto avvenga per le pellicole, tuttavia, le «registrazioni» nei materiali fotorifrangenti possono essere cancellate: le immagini possono essere memorizzate od obliterate ad arbitrio o secondo uno schema prefissato.

Grazie alla loro sensibilità e robustezza e alle loro caratteristiche ottiche uniche, i materiali fotorifrangenti potrebbero essere impiegati nella fabbricazione di elementi per l'elaborazione di dati nei calcolatori ottici. In teoria questi dispositivi consentirebbero ai calcolatori ottici di elaborare l'informazione a velocità molto più elevata dei loro corrispondenti elettronici. Impiegando materiali fotorifrangenti, i ricercatori hanno già costruito l'analogo ottico del transistore: se due fasci laser interagiscono all'interno di un materiale fotorifrangente, uno dei fasci può controllare, commutare o amplificare l'altro. I materiali fotorifrangenti costituiscono anche la base di dispositivi che mettono in risalto i contorni delle immagini, che connettono reti di laser e che memorizzano immagini tridimensionali.

Poiché le proprietà ottiche dei materiali fotorifrangenti possono essere modificate dalla stessa luce che li attraversa, questi materiali sono considerati mezzi ottici non lineari. Nei mezzi ottici lineari, come le lenti, i prismi e i filtri polarizzatori, i fasci luminosi passano semplicemente attraverso il materiale, senza modificarne in alcun modo le caratteristiche.

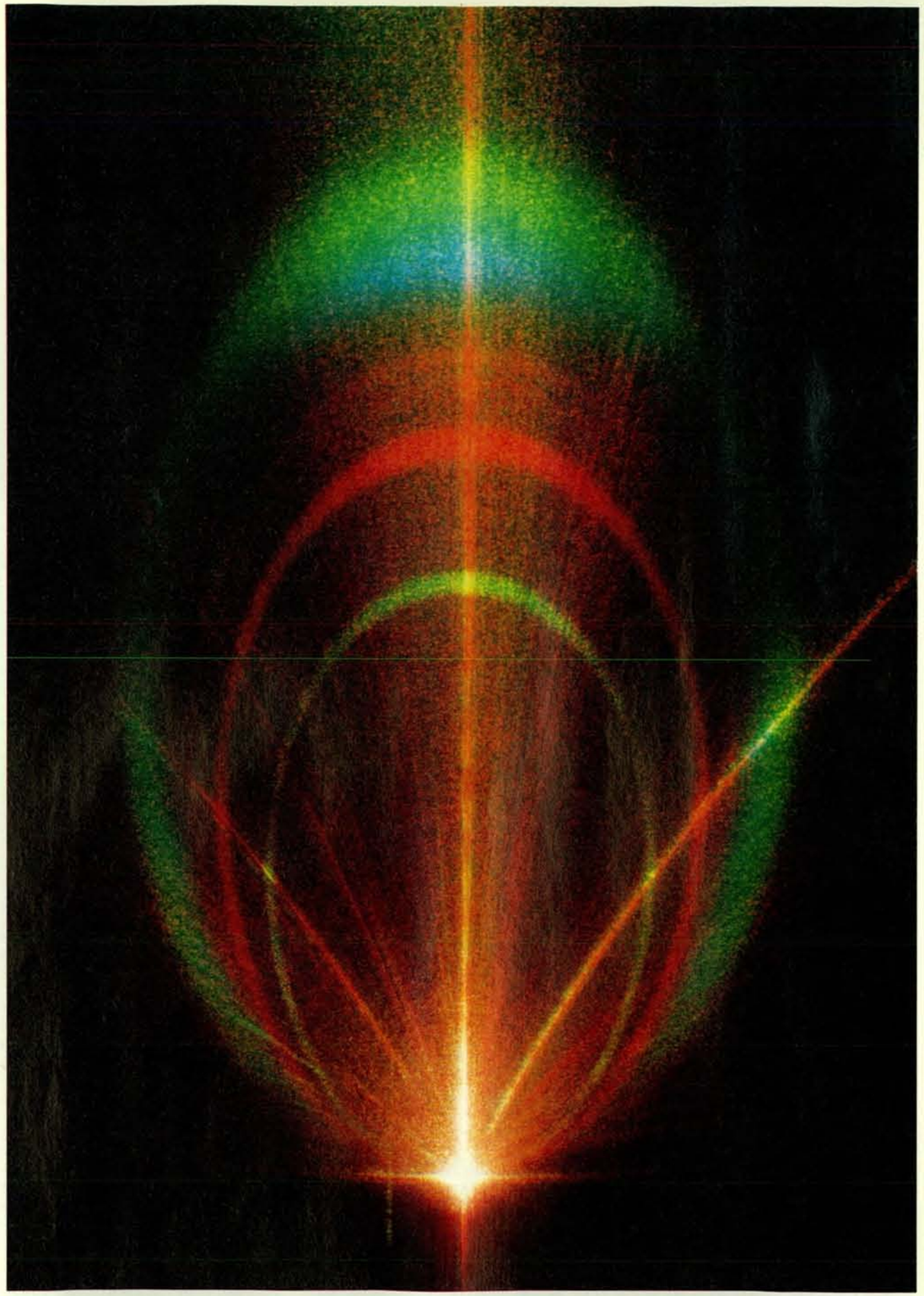
Il fenomeno della fotorifrazione è strettamente correlato con un altro fenomeno non lineare, il cosiddetto effetto fotocromatico. Colpendo con radiazione luminosa un materiale fotocromatico, si può far variare la quantità di luce assorbita dal mezzo. Certi modelli di occhiali da sole contengono materiali fotocromatici che si scuriscono al sole e si schiariscono al buio.

La luce che colpisce un materiale fotorifrangente influisce sulla velocità con cui essa stessa si propaga all'interno del materiale; più precisamente, la fotorifrazione è il fenomeno per cui la luce modifica l'indice di rifrazione di un materiale. (L'indice di rifrazione è il rapporto tra la velocità della luce nel vuoto e quella nel materiale.)

Quando vengono colpiti da una luce abbastanza intensa, quasi tutti i materiali trasparenti modificano il proprio indice di rifrazione. La luce è un'onda elettromagnetica in movimento e l'intensità del campo elettrico a essa associato è proporzionale alla radice quadrata dell'intensità della luce. Per esempio un fascio ottico con un'intensità di 100 milioni di watt al centimetro quadrato corrisponde a un campo elettrico di circa 100 000 volt al centimetro. Quando un materiale trasparente è colpito da una luce di questa intensità, le posizioni dei suoi atomi vengono sconvolte e l'indice di rifrazione varia di alcune parti su un milione. In seguito a questo cambiamento il materiale può deviare la luce, comportandosi come un prisma o una lente.

Di solito, tuttavia, il termine «fotorifrangente» è riservato a materiali il cui indice di rifrazione si modifica anche se la luce incidente è di bassa intensità. Nei materiali fotorifrangenti fasci luminosi aventi un'intensità di un solo millesimo di watt al centimetro quadrato sono in grado di alterare la disposizione degli atomi del cristallo, modificando l'indice di rifrazione anche di alcune parti su 10 000. Inoltre, a differenza di gran parte dei materiali trasparenti, questa alterazione dei cristalli fotorifrangenti è semipermanente: se, dopo aver subito l'alterazione, un cristallo viene isolato da tutte le sorgenti luminose, la variazione dell'indice di rifrazione può permanere, a seconda del materiale, da qualche millisecondo ad alcuni anni. È quindi possibile immagazzinare informazioni in un cristallo sotto forma di immagini.

Come può un debole raggio di luce causare una variazione così grande del-



Un fascio laser incidente su un cristallo di titanato di bario di cinque millimetri diffonde a ventaglio per effetto di fotorifrazione.

l'indice di rifrazione di un cristallo? Verso la fine degli anni sessanta F. S. Chen dei Bell Laboratories propose il modello fondamentale della fotorifrazione. Proprio come una sola formica può spostare un gran mucchio di sabbia trasportando un granello alla volta, così un debole fascio luminoso può creare gradualmente un forte campo elettrico spostando le cariche elettriche una alla volta. Nei cristalli fotorifrangenti le cariche diffondono allontanandosi dalle regioni illuminate e accumulandosi in quelle oscure. Questo progressivo spostamento di cariche fa aumentare il campo elettrico in seno al cristallo, che può raggiungere anche intensità di 10 000 volt al centimetro. Il campo elettrico distorce leggermente (circa dello 0,01 per cento) il reticolo cristallino e quindi modifica l'indice di rifrazione.

A quanto pare queste cariche elettriche sono prodotte da imperfezioni del reticolo cristallino del materiale. Può trattarsi di difetti meccanici della struttura reticolare (assenza di atomi in certe posizioni del reticolo), di droganti per sostituzione (presenza di atomi estranei in posizioni del reticolo), oppure di droganti interstiziali (atomi estranei incuneati fra gli atomi originari del reticolo). Bastano pochissimi difetti di questo genere, in concentrazioni di alcune parti per milione, per dar luogo al fenomeno della fotorifrazione.

Ogni difetto del cristallo può essere fonte di una carica in soprannumero: può trattarsi sia di elettroni (cariche negative), sia di buche (regioni di carica positiva), a seconda del cristallo considerato. Al buio queste cariche sono bloccate, mentre in presenza di luce sono libere di vagare nel cristallo fino a che, prima o poi, vengono di nuovo immobilizzate. Se la luce colpisce le cariche di una zona del cristallo, esse se ne allontanano e si accumulano nella regione buia, proprio come gli scarafaggi si rifugiano sotto i mobili per poter sfuggire alla luce.

Quando una carica si sposta nel cristallo, si lascia dietro una carica immobile di segno opposto. Nella zona situata tra queste due cariche di segno diverso il campo elettrico ha la massima intensità e il reticolo cristallino subisce la massima distorsione. Un fascio di luce che attraversi questa zona trova un indice di rifrazione diverso da quello delle zone del cristallo non modificate.

Il tempo che la luce impiega per ridistribuire le cariche nel cristallo dipende dalla sua intensità e anche dalla velocità con cui le cariche si spostano nel cristallo. Per generare il medesimo campo elettrico una luce debole impiega più tempo di una luce forte. Con una luce di bassa intensità (circa 0,01 watt al centimetro quadrato) possono essere necessari alcuni minuti perché le cariche raggiungano la configurazione di equilibrio, mentre con una luce d'intensità elevata (circa un miliardo di watt al centimetro quadrato) il tempo di risposta può essere inferiore al nanosecondo. Come una pellicola fotografica, anche un cristallo fotorifrangente ha bisogno di una certa quantità di luce per completare l'«esposizione».

Se il reticolo cristallino non possiede una proprietà chiamata simmetria per inversione, il cambiamento dell'indice di rifrazione è direttamente proporzionale all'intensità del campo elettrico. Quest'ultimo permane a lungo nel cristallo dopo che l'illuminazione è cessata, proprio come il mucchio di sabbia rimane a lungo nella nuova posizione dopo che le formiche se ne sono andate.

Una delle conseguenze più utili della fotorifrazione è lo scambio di energia tra due fasci laser, un fenomeno noto anche come accoppiamento tra fasci. Due fasci laser della stessa frequenza, intersecandosi, interferiscono e producono una figura immobile costituita da zone chiare e scure, o meglio uno schema di illuminazione la cui intensità varia sinusoidalmente con la posizione nel cristallo. Se questa figura di interferenza si forma all'interno di un cristallo fotorifrangente, le cariche elettriche si spostano e generano un campo elettrico la cui intensità varia pure in modo sinusoidale. Il campo risultante a sua volta distorce il reticolo cristallino in modo periodico, provocando alterazioni dell'indice di rifrazione. In definitiva nel cristallo si forma un «reticolo dell'indice di rifrazione», detto anche ologramma tridimensionale dell'indice di rifrazione.

Il campo elettrico e il reticolo dell'indice di rifrazione hanno la stessa periodicità della figura di interferenza, ma sono traslati nello spazio di un quarto di periodo rispetto alla luce incidente. Lo sfasamento di 90 gradi è la configurazione che più favorisce lo scambio di energia tra i due fasci laser di partenza.

Una volta che nel cristallo si sia formato il reticolo dell'indice di rifrazione, la luce di uno dei fasci viene in parte deviata, o diffratta, dal reticolo in direzione dell'altro fascio (e viceversa). Quindi i due fasci deviati interferiscono con i fasci incidenti, in un caso in modo costruttivo (in fase) e nell'altro in modo distruttivo (in opposizione di fase). Nel caso dell'interferenza costruttiva, le creste delle onde luminose di uno dei fasci deviati si combinano con le creste di uno dei fasci incidenti e quindi i fasci si rinforzano a vicenda; nel caso dell'interferenza distruttiva, le creste delle onde dell'altro fascio deviato si combinano con i ventri del secondo fascio incidente e le onde luminose s'indeboliscono reciprocamente. Il fascio generato dall'interferenza costruttiva emerge dal cristallo più intenso di quanto vi fosse entrato, mentre quello generato dall'interferenza distruttiva emerge indebolito. Quale dei due fasci guadagni e quale perda energia dipende dall'orientazione del cristallo e dalla natura dei portatori di carica (elettroni o buche).

Nei materiali fotorifrangenti si manifesta l'accoppiamento dei due fasci perché la figura di interferenza della luce e il reticolo dell'indice di rifrazione sono traslati nello spazio. Nella maggior parte dei materiali non lineari, tuttavia, l'accoppiamento dei fasci non si osserva perché essi rispondono «localmente» ai fasci di luce (per esempio gli orbitali atomici sono deformati dagli intensi campi elettrici dei fasci laser). Quindi in questi materiali non lineari la figura di interferenza e il reticolo corrispondono perfettamente. La luce deviata dal reticolo interferisce in modo identico con ciascuno dei fasci non deviati. Pertanto i fasci si scambiano la stessa quantità di energia e nessuno dei due aumenta d'intensità.

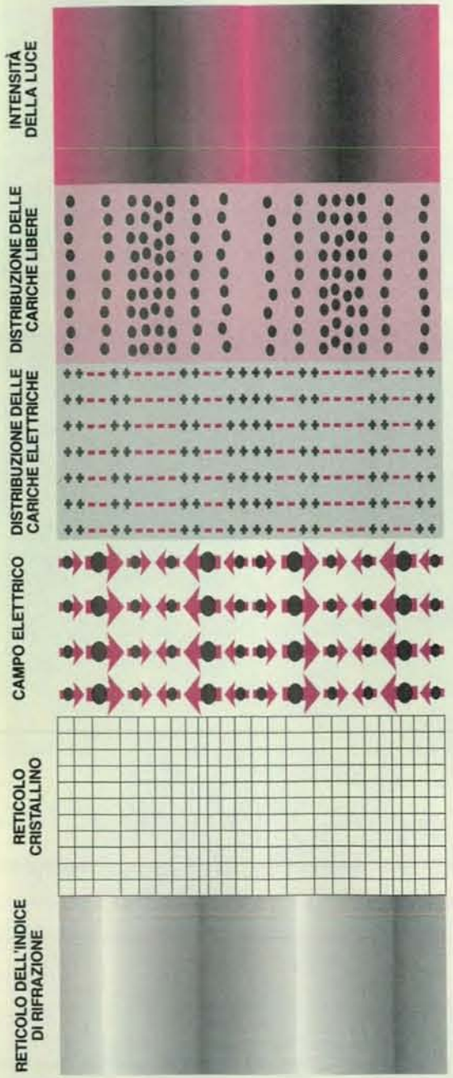
Per esaltare il fenomeno della fotorifrazione, gli studiosi hanno cominciato a controllare il flusso delle cariche in seno al materiale. In un cristallo questo flusso è regolato da due meccanismi, la diffusione e la deriva, che sono analoghi alla diffusione e alla deriva del fumo emesso da un tizzone. Di per sé il fumo tende a diffondere nelle zone in cui la sua densità è bassa; se però spira una leggera brezza, esso si sposta in una direzione particolare. Le particelle di fumo si comportano allo stesso modo delle cariche mobili di un materiale fotorifrangente: anch'esse tendono a migrare verso le regioni a bassa densità di carica e vanno alla deriva in risposta a un eventuale campo elettrico.

La semplice diffusione delle cariche dalle zone illuminate alle zone oscure del cristallo non genera il campo elettrico più intenso possibile. Nel 1981 Jean-Pierre Huignard e Abdellatif Marrakchi, dei Laboratori Thomson-CSF a Orsay, presso Parigi, applicarono un campo elettrico esterno a un cristallo fotorifrangente per generare un campo, variabile nello spazio, più intenso di quello prodotto dalla sola diffusione. Il campo elettrico applicato, tuttavia, faceva perdere al reticolo dell'indice di rifrazione lo sfasamento ottimale di un quarto di ciclo.

Per impedire questo fenomeno, Sergej I. Stepanov e Mikhail P. Petrov dell'Istituto di fisica tecnica «A. F. Ioffe» dell'Accademia delle scienze dell'Unione Sovietica, a Leningrado, idearono una tecnica assai ingegnosa. Applicando un campo elettrico esterno rapidamente alternantesi, le cariche si spostano di preferenza in una certa direzione durante la prima metà del ciclo del campo applicato e nella direzione opposta durante la seconda metà. Le cose vanno un po' come se due persone soffiassero a turno da direzioni opposte su un tizzone. La distribuzione del fumo è nello stesso tempo intensificata e più diffusa nello spazio, ma la sua posizione media è la stessa che avrebbe se nessuno soffiasse. All'interno dei cristalli fotorifrangenti questo processo genera un campo elettrico più intenso di quello prodotto dalla sola diffusione, e il reticolo dell'indice di rifrazione presenta lo stesso sfasamento medio di un quarto di ciclo che avrebbe in assenza di campo esterno.

I ricercatori hanno impiegato questa tecnica per migliorare l'efficienza dell'accoppiamento di due fasci e per rafforzare la fotorifrazione «a ventaglio» del fascio (o, in termini tecnici, diffusione in avanti stimolata per fotorifrazione). Scoperta negli anni settanta, la fotorifrazione a ventaglio è uno dei fenomeni ottici non lineari più sconcertanti. La si può osservare per esempio quando un debole fascio, del diametro di una matita, prodotto da un laser a elio-neon illumina un cristallo di titanato di bario ($BaTiO_3$). Dapprima il fascio attraversa il cristallo indisturbato, ma dopo circa un secondo (il tempo dipende dall'intensità della luce) comincia a diffondere nel cristallo flettendosi da un lato e dividendosi in molti raggi che si aprono in un ampio ventaglio di luce, da cui il nome del fenomeno. A seconda della scelta del cristallo fotorifrangente, la luce uscente può possedere diverse componenti di polarizzazione e la sua sezione può assu-

COME LA LUCE ALTERA LE PROPRIETÀ OTTICHE DI UN CRISTALLO FOTORIFRANGENTE



1. Due fasci laser interferiscono all'interno di un cristallo fotorifrangente generando una figura di zone chiare e scure.

2. Gli elettroni liberi si allontanano dalle zone chiare del cristallo.

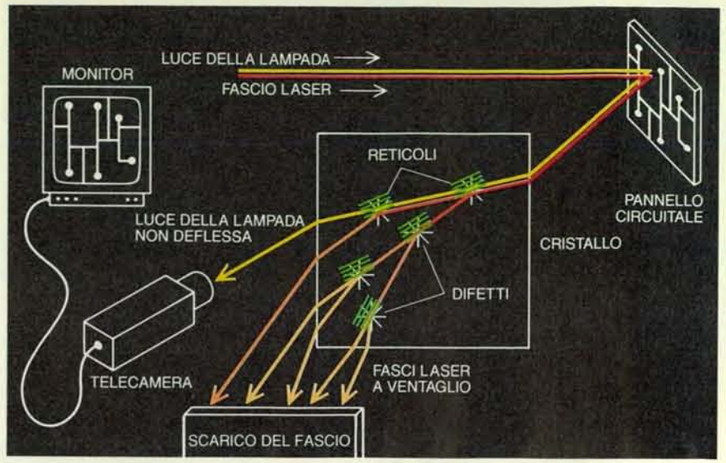
3. Gli elettroni liberi si accumulano nelle zone scure e si formano zone di carica positiva.

4. Tra le zone di carica positiva e negativa si forma un campo elettrico.

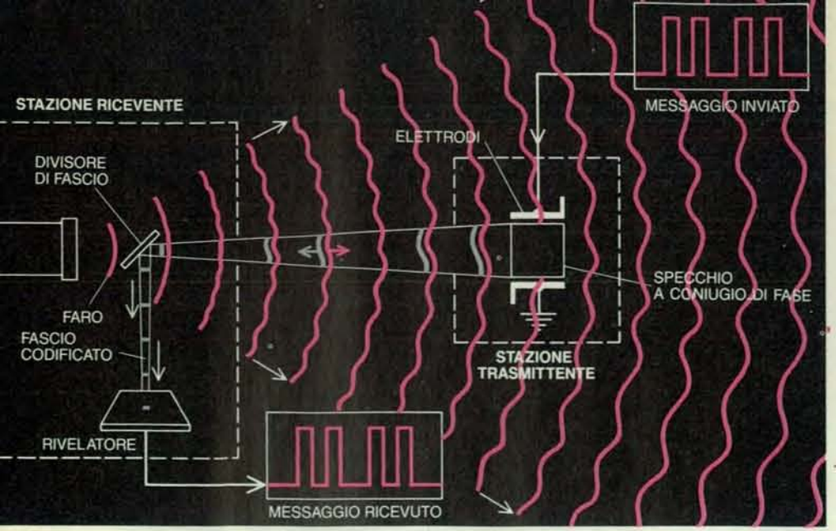
5. Il campo elettrico distorce il reticolo cristallino.

6. A causa della distorsione la luce si propaga a velocità più alta in certe zone e più bassa in altre. L'indice di rifrazione subisce cioè una variazione periodica. Il reticolo dell'indice di rifrazione è spostato nello spazio di un quarto di periodo rispetto alla figura d'interferenza luminosa.

Nell'illustrazione è raffigurata schematicamente la successione di processi fisici mediante i quali la luce modifica le proprietà ottiche di un cristallo di materiale fotorifrangente.



L'escissore ottico coerente diffonde l'intensa luce di un laser, ma trasmette la debole luce di una lampada. Nell'illustrazione l'escissore protegge una telecamera dai raggi laser che altrimenti la danneggerebbero. Il fascio laser, composto da luce monocromatica, viene diffuso dai difetti del cristallo fotorifrangente e poi interferisce con se stesso. La figura d'interferenza genera un reticolo dell'indice di rifrazione. Di conseguenza gran parte della luce laser viene diffusa verso un lato del cristallo, con un effetto detto di fotorifrazione a ventaglio. Poiché la luce della lampada è composta di molte frequenze, non si genera né una figura d'interferenza né un reticolo, e quindi quasi tutta la luce attraversa il cristallo.



Uno specchio a coniugio di fase costruito con un cristallo fotorifrangente consente le comunicazioni in atmosfera. La stazione ricevente emette un fascio di luce che, propagandosi, si allarga e si distorce. Alla stazione trasmittente un cristallo fotorifrangente riflette o trasmette la luce che lo colpisce a seconda della tensione applicata agli elettrodi. Poiché il cristallo funge da specchio a coniugio di fase, la luce riflessa subisce un'inversione nel tempo. Un messaggio può essere codificato riflettendo e trasmettendo in alternanza la luce. Quando il fascio contenente il messaggio, invertito nel tempo, interagisce con l'atmosfera, tutte le distorsioni vengono rimosse e il messaggio può essere decodificato alla stazione ricevente.

mere la forma di una pittoresca famiglia di figure ellittiche.

La fotorifrazione a ventaglio del fascio ha origine da uno scambio di energia tra il fascio luminoso incidente e la luce diffusa in modo casuale dai difetti del cristallo. La luce diffusa da un difetto interferisce con il fascio non diffuso formando un reticolo dell'indice di rifrazione a causa dell'effetto di fotorifrazione. Quando ciò sia avvenuto, la luce non diffusa può trasferire altra energia al fascio diffuso. Il meccanismo di scambio energetico risulta anisotropo e quindi i fasci luminosi diffusi vengono intensificati prevalentemente in una gamma finita di angoli. Il fascio diffuso e intensificato subisce poi nuove diffusioni casuali dovute ad altri difetti del cristallo e il processo si ripete. Di conseguenza dal fascio originale esce a ventaglio un gran numero di fasci diffusi. Variando l'orientazione del cristallo e cambiando il punto di incidenza del fascio laser si può fare in modo che il fascio fotorifratto a ventaglio subisca una piccola flessione (cioè curvi appena percettibilmente attraversando il cristallo) oppure si fletta di un angolo piuttosto grande (e colpisca allora uno spigolo lontano del cristallo).

La fotorifrazione a ventaglio del fascio e altri fenomeni di fotorifrazione sono stati sfruttati in applicazioni quali escissori di fasci coerenti, elementi ottici di interconnessione, specchi a coniugio di fase, filtri delle variazioni e sistemi per intensificare i contorni delle immagini.

L'escissore ottico coerente, descritto per la prima volta nel 1985 da Mark Cronin-Golomb e Amnon Yariv del California Institute of Technology, può in teoria filtrare, eliminandola, la luce che diffonde da un fascio laser d'intensità elevata. L'escissore è in grado di assicurare la protezione di delicati rivelatori ottici, cosa che può dimostrarsi necessaria nel controllo dei laser di grande potenza impiegati in certi processi industriali come la ricottura e la saldatura. L'escissore può anche impedire il danneggiamento delle telecamere (*si veda l'illustrazione a pagina 59*).

Il componente fondamentale dell'escissore ottico coerente è un cristallo fotorifrangente collocato di fronte al rivelatore o all'obiettivo della telecamera. Il fascio intenso può essere stornato dal rivelatore grazie alla fotorifrazione a ventaglio, che in sostanza fa deviare di lato la luce laser coerente che attraversa il cristallo. Il rivelatore o la telecamera può vedere l'oggetto anche in presenza di questa luce intensa perché l'illuminazione di fondo dell'ambiente, diffusa dal resto dell'oggetto, attraversa il cristallo quasi senza alterazioni.

La fotorifrazione a ventaglio ha un ruolo importante anche negli specchi a coniugio di fase costruiti con materiali fotorifrangenti. Questi specchi godono della peculiare proprietà che un fascio ottico da essi «riflesso» si propaga all'indietro nello spazio esattamente come se il tempo fosse stato invertito. Grazie a questa proprietà, gli specchi a coniugio di fase trovano moltissime applicazioni nei settori delle comunicazioni ottiche, dei laser di grande potenza e dell'elaborazione ottica. Possono per esempio essere incorporati in sistemi che correggono le aberrazioni indesiderate subite talvolta dai fasci laser nel propagarsi attraverso mezzi distorcenti o potenti amplificatori laser (si veda l'articolo *Le applicazioni del coniugio ottico di fase* di David M. Pepper in «Le Scienze» n. 211, marzo 1986).

Nel 1977 Robert W. Hellwarth della University of Southern California propose la configurazione base per gli specchi a coniugio di fase e due anni dopo Sergej G. Odulov e uno di noi (Kukhtarev) e, indipendentemente, Huignard e collaboratori costruirono uno specchio a coniugio di fase che impiegava materiali fotorifrangenti. Nel 1982 uno di noi (Feinberg) scoprì fortuitamente una classe di specchi a coniugio di fase oggi utilizzata da molti ricercatori. Feinberg aveva diretto tre fasci laser su un cristallo di titanato di bario: un fascio conteneva le onde luminose di cui si desiderava la replica invertita nel tempo; gli altri due fasci di «pompaggio» erano necessari per formare lo specchio a coniugio di fase (almeno così ritenevano allora Feinberg e colleghi). Nel controllare il dispositivo sperimentale Feinberg bloccò i fasci di pompaggio per accertarsi che il presunto fascio invertito nel tempo non derivasse da una semplice riflessione su una faccia del cristallo. In un primo momento il fascio invertito nel tempo spariva, ma dopo qualche istante inopinatamente ricompariva. Feinberg aveva scoperto uno specchio a coniugio di fase che richiedeva un solo fascio. Questo elegante dispositivo non è che un esempio di una classe più generale di specchi a coniugio di fase autopompati che ebbe i suoi primi studiosi in Jeffrey O. White, Mark Cronin-Golomb, Baruch Fischer e Amnon Yariv del California Institute of Technology.

Benché gli specchi a coniugio di fase si possano costruire con molti tipi di materiali ottici non lineari, gli elementi fotorifrangenti presentano molti e chiari vantaggi. In primo luogo gli specchi richiedono un solo fascio d'ingresso, cioè proprio il fascio di cui si cerca la replica coniugata in fase: si ottiene in questo modo il cosiddetto coniugatore di fase autopompato. In secondo luogo, il processo che produce il fascio coniugato in fase può essere innescato con potenze e intensità del laser molto basse.

Perché questa predilezione della natura per il fascio coniugato in fase? Una risposta parziale a questo interrogativo posto da Hellwarth può essere, almeno nel caso del titanato di bario, quella avanzata da Kenneth MacDonald, allora alla University of Southern California, e da uno di noi (Feinberg). Dopo un po' di tempo la fotorifrazione a ventaglio fa sì che la luce laser incidente si fletta di preferenza verso un lato del cristallo. Se il fascio incidente e il cristallo sono disposti in modo che il fascio fotorifratto a ventaglio venga spinto verso l'angolo opposto del cristallo, questo fascio subisce due riflessioni interne, cioè in pratica si ripiega su se stesso. Il fascio riflesso subisce ancora fotorifrazione a ventaglio all'indietro, nella direzione del fascio incidente. Di tutti i fasci diffusi all'interno del cristallo, il fascio invertito nel tempo è quello che, grazie alla sua traiettoria retrograda, acquista più energia. Questo fenomeno di inversione del fascio può raggiungere un'efficienza molto elevata, tanto che il fascio coniugato in fase può contenere anche il 60 per cento della potenza del fascio incidente.

Uno di noi (Pepper) ha aggiunto un ulteriore tocco a questo nuovo specchio a coniugio di fase. Collegando a un cristallo fotorifrangente degli elettrodi, Pepper gli ha applicato un campo elettrico variabile nel tempo. Colpendo il cristallo, il fascio laser non solo viene invertito nel tempo, ma può essere anche modulato a impulsi, come se il dispositivo fosse uno specchio munito di otturatore. In questo modo si possono trasmettere informazioni a impulsi dallo specchio a coniugio di fase alla sorgente laser: l'inversione temporale del fascio assicura che i due punti in comunicazione rimangano «agganciati» tra loro. Questo schema può essere usato per istituire un canale di comunicazione tra due satelliti o per trasmettere informazioni da un sensore remoto situato all'estremità di un collegamento a fibra ottica (*si veda l'illustrazione nella pagina a fronte*).

Un altro dispositivo in cui si sfrutta un fenomeno di scambio energetico è il cosiddetto «filtro delle variazioni», che mette in evidenza tutto ciò che cambia in una scena molto complessa. Dispositivi di questo tipo possono evidenziare aerei in volo su uno sfondo di edifici, un atleta che si tuffa in una piscina tranquilla o batteri natanti su uno sfondo di alghe immobili.

Cronin-Golomb (ora alla Tufts University) ha allestito un tipo di filtro delle variazioni in cui due fasci illuminano un cristallo fotorifrangente. Il primo fascio laser contiene la codifica tridimensionale di un'immagine. Il cristallo viene orientato in modo che il fascio contenente l'immagine trasferisca gran parte della propria energia al secondo fascio. Dopo aver attraversato il cristallo il fascio contenente l'immagine è quasi del tutto oscurato. Quando però qualcosa cambia in questo fascio, il meccanismo di scambio energetico subisce una momentanea perturbazione e la porzione di fascio in cui si è verificato il cambiamento attraversa il cristallo e può essere osservata su uno schermo. Quando il movimento cessa il fascio contenente l'immagine diventa nuovamente oscuro dopo l'attraversamento del cristallo.

I cristalli fotorifrangenti possono essere usati anche in certi dispositivi che intensificano i contorni di un'immagine. Quest'ultima viene codificata in un fascio «oggetto», che è proiettato in un cristallo fotorifrangente insieme con un fascio di «riferimento». I due fasci interferiscono all'interno del cristallo e producono un ologramma dell'immagine originale, la quale può essere poi recuperata da un terzo fascio, di «lettura», che si propaga in verso opposto rispetto al fascio di riferimento. Se il fascio oggetto è relativamente debole, l'immagine ricostruita è una replica fedele di quella di partenza. Se invece il fascio oggetto è più intenso degli altri due, verranno messi in risalto i contorni dell'immagine ricostruita.

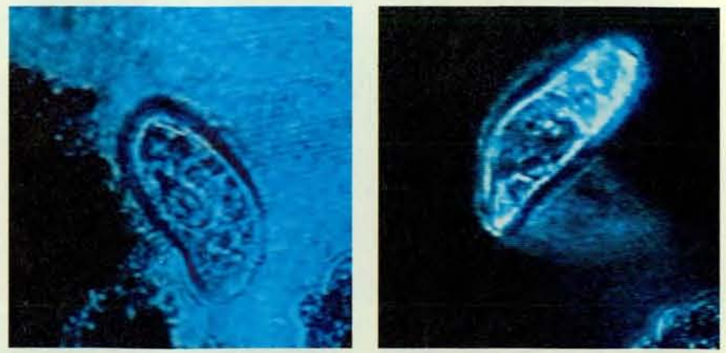
L'intensità del fascio oggetto varia localmente nel cristallo perché il fascio contiene un'immagine. Perciò la sua intensità coincide con quella del fascio di riferimento in corrispondenza dei contorni perché ogni contorno contiene tutta la gamma delle intensità, dal chiaro allo scuro. Dove le intensità dei due fasci coincidono, la figura d'interferenza ottica in quella particolare regione del cristallo ha la massima modulazione da chiaro a scuro a chiaro e così via. Questa figura d'interferenza così contrastata produce un netto reticolo dell'indice di rifrazione. Quando il fascio di lettura interagisce con il reticolo, viene deviato, o diffratto, con la massima efficienza in quella posizione, e quindi l'immagine ricostruita emerge con tutti i suoi contorni intensificati.

Siccome i cristalli fotorifrangenti possono fungere sia da accoppiatori di energia sia da specchi a coniugio di fase, sono particolarmente utili per collegamenti ottici ristrutturabili e per l'agganciamento in frequenza dei laser. I cristalli fotorifrangenti possono trasmettere informazioni da un elemento ottico a un altro, per esempio da una fibra ottica a un dispositivo per l'elaborazione di dati, senza bisogno di complicati elementi ottici o connessioni elettroniche. Questo sistema ottico di trasmissione può anche costringere due (o più) laser separati ad «agganciarsi» tra loro, in modo che oscillino alla stessa identica frequenza ottica; in questo modo i due laser si comportano nel complesso come un solo laser più grande.

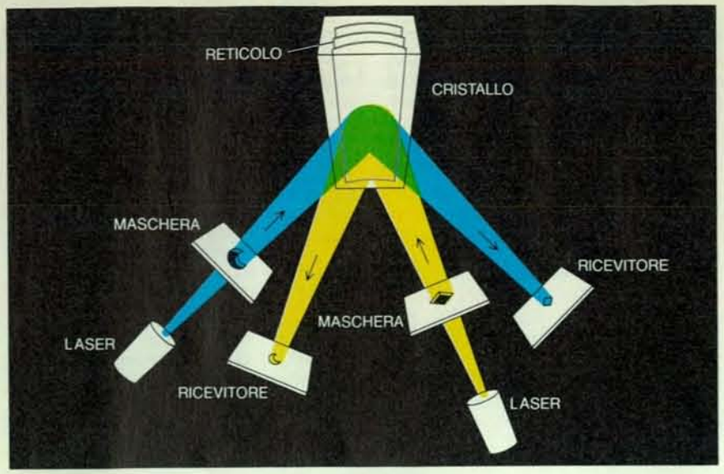
Come fanno due o più fasci di luce tra loro non coerenti a connettersi in un cristallo fotorifrangente? Quando i diversi fasci illuminano un cristallo fotorifrangente, ciascuno produce la propria moltitudine di fasci diffusi e di reticoli casuali dell'indice di rifrazione (ologrammi) in seno al cristallo. Se uno degli ologrammi di un certo fascio coincide esattamente con un ologramma dell'altro fascio, allora l'intensità di quest'ultimo cresce più rapidamente di quella degli altri. Ne segue che i due fasci si trovano connessi l'uno all'altro. Un dispositivo di questo tipo è chiamato specchio a coniugio di fase a pompaggio reciproco o a pompaggio doppio.

Il dispositivo, messo a punto per la prima volta nel 1987 da Fischer, che attualmente lavora al Technion-Israel Institute of Technology di Haifa, e collaboratori, da Robert W. Eason e A. M. C. Smout dell'Università dell'Essex in Gran Bretagna e più tardi da Mark D. Ewbank del Rockwell International Science Center di Thousand Oaks, in California, può connettere due fasci provenienti da due direzioni qualunque. Se i due fasci contengono immagini, queste vengono trasformate l'una nell'altra durante il passaggio dei fasci nel cristallo (*si veda l'illustrazione a pagina 62*). Se un fascio è modulato a impulsi a frequenza elevata, l'informazione temporale viene ritrasmessa all'indietro all'altro fascio.

Se i fasci provengono da due laser oscillanti a frequenze leggermente diverse, uno dei fasci viene deviato nella direzione approssimativamente coniuga-



Un paramecio si trova tra alghe e detriti (*a sinistra*), ma quando la scena viene osservata attraverso un filtro delle variazioni, si vede solo il microrganismo che nuota (*a destra*). Il filtro delle variazioni, progettato da R. M. Pierce, R. Cudney, G. D. Bacher e J. Feinberg, è in grado di evidenziare qualunque oggetto in moto rispetto a uno sfondo immobile.



Un cristallo fotorifrangente accoppia due fasci laser tramite un meccanismo che può essere sfruttato nei sistemi di comunicazione, nei calcolatori ottici o in altre reti a laser.

ta, ossia invertita nel tempo, del fascio proveniente dall'altro laser e viceversa. Se i due laser sono connessi otticamente da un cristallo, in condizioni opportune possono essere fatti oscillare in modo coerente, ottenendo così l'agganciamento delle frequenze. Sfruttando questa configurazione, i ricercatori sperano di poter combinare in modo efficiente migliaia di laser a semiconduttore e di formare una potentissima sorgente di luce coerente.

Per costruire componenti e dispositivi ottici migliori, gli studiosi cercano tra i molti tipi diversi di materiali fotorifrangenti i cristalli più efficienti e affidabili. Le proprietà ottiche, elettriche e strutturali dei materiali fotorifrangenti variano moltissimo: si hanno così il titanato di bario, che è isolante, l'arseniuro di gallio, semiconduttore, e il composto organico 2-ciclottilammino-5--nitropiridina drogato con 7,7,8,8-tetracianochinodimetano.

Questi materiali in apparenza così diversi manifestano nondimeno effetti di fotorifrazione simili. Hanno tutti un reticolo cristallino relativamente facile da distorcere e contengono tutti difetti che fungono da sorgenti di portatori di carica e da trappole per le cariche. Sottoposte a un campo elettrico, tuttavia, le cariche si muovono con velocità circa 10 000 volte maggiore nell'arseniuro di gallio che nel titanato di bario. Pertanto, a parità d'intensità della luce incidente, la fotorifrazione si sviluppa molto più rapidamente nel primo materiale che nel secondo. Tuttavia il reticolo dell'indice di rifrazione risultante è più netto nel titanato di bario che nell'arseniuro di gallio. Anche le proprietà ottiche dei cristalli variano molto da un materiale all'altro: l'arseniuro di gallio e gli altri semiconduttori sono in genere fotosensibili nella zona del vicino infrarosso dello spettro ottico, mentre gli isolanti e i composti organici sono per lo più fotosensibili nello spettro visibile.

Una miriade di esperimenti attualmente in corso è mirata a identificare e possibilmente a controllare i difetti che danno luogo al fenomeno della fotorifrazione nei vari cristalli. Per esempio non è stata ancora identificata la causa della fotorifrazione nel titanato di bario: molti ricercatori attribuiscono il fenomeno ai diversi stati di ionizzazione dei metalli di transizione, ferro, cobalto e manganese, presenti in tracce. Anche le vacanze di atomi di ossigeno nel reticolo possono contribuire alla formazione di portatori di carica. Nell'arseniuro di gallio, viceversa, il fenomeno di fotorifrazione è attribuito a un difetto intrinseco del reticolo, che si pensa derivi dall'effetto combinato della sostituzione nel reticolo di un atomo di gallio con un atomo di arsenico e dell'introduzione di un atomo di arsenico in più nella stessa cella con formazione del centro EL2.

A prescindere dalla natura dei difetti del cristallo, le proprietà di gran parte dei materiali fotorifrangenti possono essere modificate drogandoli con impurezze oppure estraendo atomi dal loro reticolo. Per esempio uno di noi (Feinberg) e Stephen Ducharme della University of Southern California hanno scoperto che quando un cristallo di titanato di bario viene riscaldato in assenza di ossigeno per rimuovere parzialmente l'ossigeno dal reticolo cristallino, le proprietà di fotorifrazione vengono alterate profondamente perché i portatori di carica predominanti non sono più le buche bensì gli elettroni.

Facendo aumentare la concentrazione dei difetti nei materiali fotorifrangenti, i ricercatori hanno scoperto che il numero delle cariche disponibili aumenta e simultaneamente cresce l'intensità del campo elettrico interno e del reticolo dell'indice di rifrazione. D'altra parte i difetti diffondono e assorbono la luce dei fasci incidenti. La concentrazione ottimale dei difetti, nel senso che fornisce un numero di cariche ragionevole senza attenuare sensibilmente la luce incidente, sembra compresa tra una e 100 parti per milione.

Forse i migliori materiali fotorifrangenti saranno quelli ricavati, con grande difficoltà, dai semiconduttori. Di recente Stephen E. Ralph, David D. Nolte e Alastair M. Glass degli AT&T Bell Laboratories hanno dimostrato la possibilità di costruire strutture combinando strati di leghe di arseniuro di gallio e alluminio dello spessore di pochi atomi. Queste strutture, chiamate super-reticoli e pozzi quantici, manifestano una fotorifrazione misurabile. Questi materiali sono oggi studiati per le loro particolari proprietà elettriche e per l'alta velocità di risposta, e i loro cristalli potrebbero anche essere impiegati in una nuova classe di unità di elaborazione integrate basate sull'ottica e sull'elettronica.

Ciò che rende i materiali fotorifrangenti tanto promettenti per i calcolatori ottici è la loro grande sensibilità alla luce, combinata con la capacità di collegare tra loro fasci emessi da laser diversi e di trasformare una forma d'informazione in un'altra. Il futuro dei materiali fotorifrangenti dipenderà dalla possibilità di specificare le loro proprietà ottiche con la stessa precisione che si è raggiunta nel caso dei semiconduttori. L'ideale sarebbe poter combinare i dispositivi fotorifrangenti con laser e rivelatori a semiconduttore in modo da ottenere un unico dispositivo compatto, in grado di elaborare simultaneamente a ogni microsecondo milioni di bit di dati, con una velocità di elaborazione complessiva dell'ordine di mille miliardi di bit al secondo.

BIBLIOGRAFIA

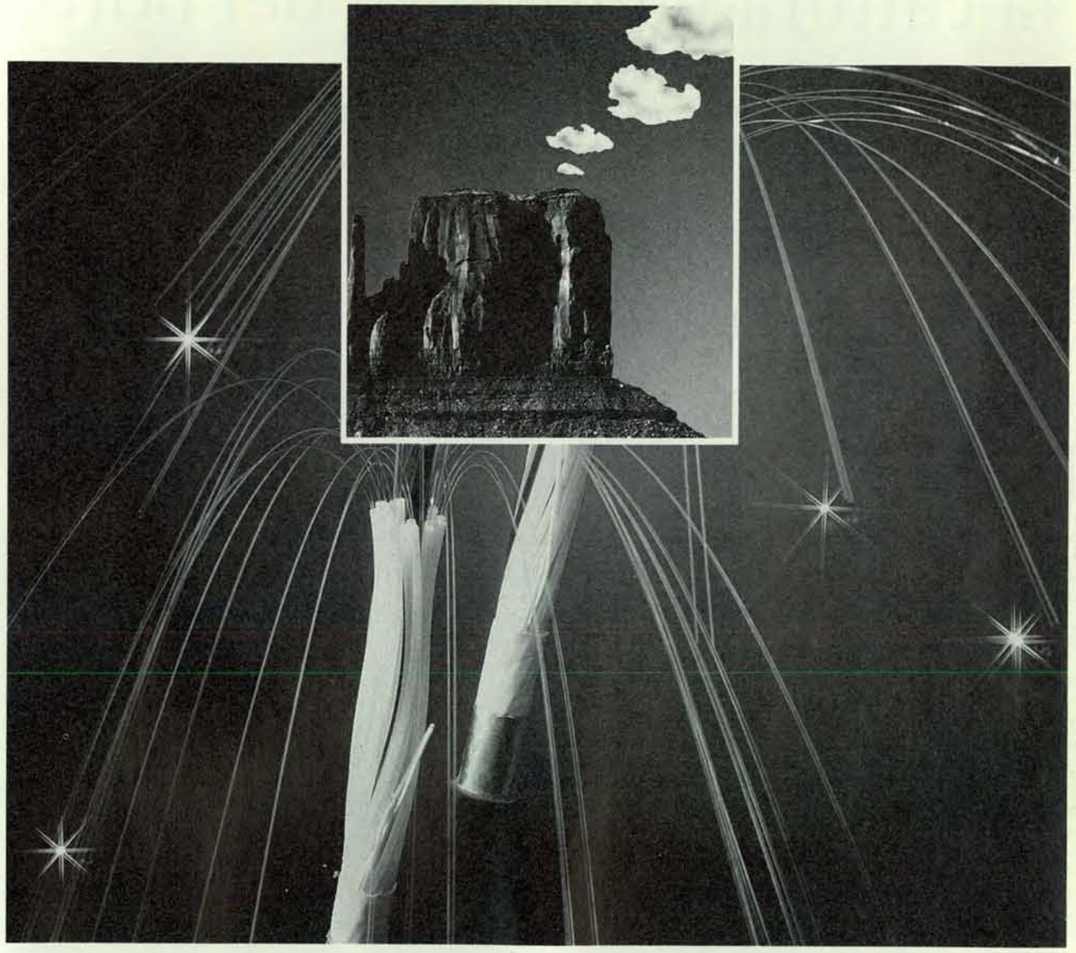
FEINBERG JACK, *Optical Phase Conjugation in Photorefractive Materials* in *Optical Phase Conjugation*, a cura di R. A. Fisher, Academic Press, 1983.

FEINBERG JACK, *Photorefractive Nonlinear Optics* in «Physics Today», 41, n. 10, ottobre 1988.

GÜNTER P. e HUIGNARD J.-P. (a cura), *Photorefractive Materials and Their Applications I & II: Topics in Applied Physics*, vol. 61 e 62, Springer-Verlag, 1988.

Nonlinear Optical Phase Conjugation, numero speciale di «IEEE Journal of Quantum Electronics» a cura di David M. Pepper, 25, n. 3, marzo 1989.

DAI SEGNALI DI FUMO
ALLA TECNOLOGIA DELLE FIBRE OTTICHE



L'INTELLIGENZA DELL'UOMO
VERSO UN MODO DI COMUNICARE.

L'evoluzione dell'uomo è evoluzione della sua capacità di comunicare.

Dai graffiti rupestri ai tam tam, ai segnali di fumo, agli impulsi elettrici, ed oggi, ai lampi di luce che corrono lungo sottili fili di vetro: è l'intelligenza dell'uomo che sviluppa avanzati sistemi di rilevamento e controllo per la trasmissione e lo scambio di informazioni.

A questa straordinaria evoluzione partecipa la Sirti che studia, progetta, realizza reti di telecomunicazioni a lunga distanza, reti urbane, reti locali, reti metropolitane, reti per telecomunicazioni ferroviarie, collegamenti in cavo sottomarino, sistemi per gestione, supervisione e controllo.

Know how tecnologico, know how organizzativo identificano la Sirti come leader dell'impiantistica delle telecomunicazioni in Italia e all'estero.

Sirti
costruisce il mondo delle telecomunicazioni

Una nuova terapia anticancro: la cattura neutronica del boro

Un trattamento in due fasi che associa l'azione di un isotopo stabile del boro con quella di neutroni di bassa energia consente di irradiare e distruggere le cellule cancerose senza danneggiare i tessuti sani circostanti

di Rolf F. Barth, Albert H. Soloway e Ralph G. Fairchild

La terapia ideale per il cancro dovrebbe consentire di distruggere le cellule tumorali senza danneggiare gravemente i tessuti sani circostanti. Per ottenere la guarigione del paziente è necessario eradicare totalmente il tumore: infatti, se le cellule cancerose non vengono distrutte fino all'ultima vi è il pericolo che il male possa ripresentarsi. Sebbene oggi i trattamenti convenzionali - chirurgia, radioterapia e chemioterapia - riescano a curare un certo numero di forme di cancro, vi sono ancora molte eccezioni. La possibilità di mettere a punto un'altra efficace terapia ha condotto gli autori di questo articolo e i loro colleghi di altri centri di ricerca a interessarsi di una nuova metodologia: la terapia con cattura neutronica del boro (BNCT, da *boron neutron capture therapy*).

Questa tecnica combina due componenti che di per sé hanno effetti molto ridotti sulle cellule normali. Il primo componente è un isotopo stabile del boro (boro 10) che può essere concentrato nelle cellule tumorali; il secondo è un fascio di neutroni di bassa energia che produce radiazione a corto raggio quando viene assorbito, o «catturato», dal boro. La combinazione di questi fattori nel sito di un tumore libera un'intensa radiazione che è in grado di distruggere i tessuti cancerosi.

La BNCT si basa sulla reazione nucleare che avviene quando il boro 10, un isotopo non radioattivo che costituisce circa il 20 per cento del boro presente in natura, viene irradiato con neutroni e li assorbe. I neutroni che esso cattura sono i cosiddetti neutroni termici, o lenti. La loro energia è così bassa (circa 0,025 elettronvolt) che essi causano danni minimi ai tessuti rispetto ad altre forme di radiazione come protoni, raggi gamma e neutroni veloci.

Quando un atomo di boro 10 cattura un neutrone si forma un isotopo instabile, il boro 11, che subisce istantaneamente fissione dando origine a nuclei di litio 7 e a particelle alfa di alta energia. Queste particelle pesanti (costituite da due protoni e due neutroni), con un'energia di 2,79 milioni di elettronvolt, costituiscono una forma di radiazione dagli effetti letali per le cellule.

Se il trattamento procede regolarmente, gli effetti distruttivi della reazione di cattura avvengono in primo luogo in quelle cellule tumorali che hanno accumulato boro 10, mentre le cellule normali con basse concentrazioni di boro vengono risparmiate. Tuttavia un fattore limitante per questo procedimento è costituito dal fatto che reazioni di cattura analoghe a quelle del boro hanno luogo anche con l'azoto e l'idrogeno nelle altre cellule dell'organismo. Fortunatamente la sezione d'urto di cattura neutronica dell'azoto e dell'idrogeno (che viene espressa in un'unità di misura detta barn, pari a 10^{-24} centimetri quadrati) è piccola rispetto a quella del boro 10, e quindi la quantità di radiazione generata nelle cellule non interessate dal trattamento con boro è del tutto trascurabile. (La sezione d'urto di cattura esprime la probabilità che un dato atomo catturi una particella in arrivo, in questo caso un neutrone.)

I tessuti sani, però, contengono grandi quantità di azoto e idrogeno (rispettivamente il 3 e il 10 per cento in peso); pertanto la dose massima di neutroni con cui può essere irradiato un tumore è determinata in ultima analisi dalla tolleranza dei tessuti circostanti alla radiazione (protoni e raggi gamma) emessa dalle reazioni di cattura neutronica dell'azoto e dell'idrogeno.

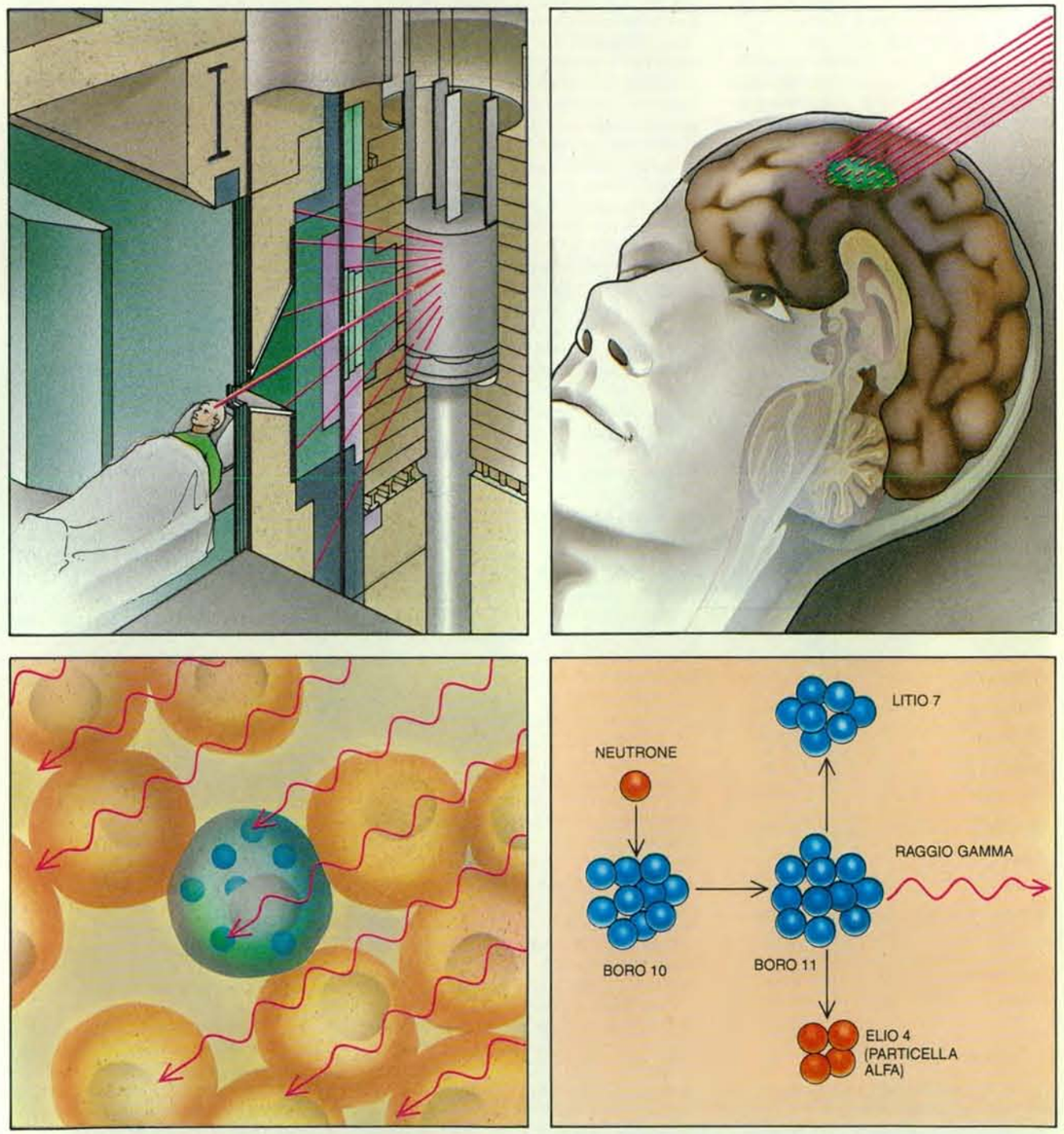
Una delle caratteristiche più interessanti della BNCT è che il tipo predominante di radiazione emessa dalle reazioni di cattura è rappresentato da particelle alfa. Si tratta di nuclei di elio, composti da due protoni e due neutroni, ma privi di elettroni orbitanti. Le particelle alfa si propagano piuttosto lentamente; nei tessuti biologici hanno un cammino della lunghezza di circa un diametro cellulare (10 micrometri) e danno origine a effetti di ionizzazione in una regione molto ristretta. Oltre a ciò, è sufficiente l'energia emessa da poche particelle alfa all'interno di una cellula tumorale per distruggerla. Queste caratteristiche fanno, in teoria, della BNCT un metodo ideale per la distruzione di cellule neoplastiche.

Le particelle alfa presentano altri vantaggi dal punto di vista biologico. Al contrario di altre forme di radiazione ionizzante, come i raggi X, non richiedono ossigeno per esplicare i loro effetti sui tessuti. Un tumore in rapida espansione è spesso scarsamente irrorato, cosicché alcune sue regioni ricevono meno ossigeno rispetto ai tessuti sani. [Questa tesi contrasta con gli studi e le sperimentazioni di altri ricercatori; si veda per esempio l'articolo *La vascolarizzazione dei tumori* di Judah Folkman in «Le Scienze» n.97, settembre 1976.] A causa di questa carenza di ossigeno, il tumore può acquistare una certa resistenza all'azione della radioterapia convenzionale. Esso rimane però sensibile alle particelle alfa perché queste, non richiedendo ossigeno, possono aggredire con la stessa energia tutte le parti di un tumore.

Un altro vantaggio delle particelle alfa è che esse sono in grado di distruggere sia le cellule neoplastiche in mitosi (ossia in corso di divisione) sia quelle che si trovano nella fase di riposo (ossia non si stanno dividendo, come spesso accade nei tumori, dove vi è un numero considerevole di cellule funzionali ma inattive). Le altre forme di radioterapia e chemioterapia tendono ad avere efficacia solo sulle cellule in mitosi.

Il concetto che sta alla base della BNCT non è certo nuovo; fu proposto per la prima volta nel 1936 da Gordon L. Locher, che allora lavorava in un reparto del Franklin Institute di Swarthmore, in Pennsylvania. Tuttavia la tecnica è stata presa in considerazione solo di recente a causa della difficoltà di introdurre selettivamente nelle cellule tumorali un numero sufficiente di atomi di boro 10 e di fare in modo che un adeguato numero di neutroni inneschi la reazione di cattura neutronica. Questi problemi non sono stati ancora del tutto risolti, ma si sono compiuti progressi in entrambi i campi. Diversi laboratori negli Stati Uniti, in Europa, in Giappone e in Australia stanno lavorando attivamente alla ricerca di una soluzione.

Si stanno studiando possibili portatori di boro quali farmaci, anticorpi monoclonali e derivati di composti naturali. Un flusso adeguato di neutroni può essere fornito da reattori nucleari e da altre fonti. La radiazione prodotta da un reattore comprende raggi gamma, neutroni veloci, neutroni epitermici (con energia compresa tra uno e 10 000 elettronvolt) e neutroni termici. Il problema che fisici e ingegneri dovranno risolvere è



La terapia con cattura neutronica del boro (BNCT) può essere eseguita in un centro dotato di un reattore nucleare o in un ospedale che disponga di una fonte alternativa di neutroni. Un fascio di neutroni epitermici penetra nel tessuto cerebrale e raggiunge la massa tumorale. Qui i neutroni di bassa energia si combinano con il boro 10 (introdotto in precedenza nelle cellule neoplastiche mediante farmaci o anticorpi) formando boro 11, che emette radiazioni (particelle alfa) in grado di distruggere il tumore.

quello di ottenere fasci costituiti soprattutto da neutroni epitermici e termici.

I primi studi clinici sulla BNCT si svolsero negli anni cinquanta e all'inizio degli anni sessanta. Fra i primi sperimentatori vi furono Lee E. Farr, James S. Robertson ed Elmer E. Stickley del Brookhaven National Laboratory e William H. Sweet, Gordon L. Brownell e uno di noi (Soloway) del Massachusetts General Hospital e del Massachusetts Institute of Technology. Le loro ricerche si concentrarono sul trattamento del glioblastoma multiforme, il più maligno fra tutti i tumori cerebrali. Esso venne prescelto, fra l'altro, perché le cellule neoplastiche che quasi inevitabilmente sfuggono alla rimozione chirurgica provocano lo sviluppo di metastasi nel tessuto cerebrale circostante, causando la morte del paziente.

La BNCT non riuscì a prolungare la sopravvivenza di questi pazienti, anzi in qualche caso si manifestarono complicazioni dovute all'irradiazione. Gli sperimentatori conclusero che i composti del boro non si concentravano né permanevano a lungo nelle cellule neoplastiche e che i neutroni termici non penetravano fino a profondità adeguata. Di conseguenza si avevano danni eccessivi ai tessuti sani e una distruzione incompleta dei tumori profondi.

Soloway e Sweet iniziarono a cercare composti del boro che si concentrassero in maniera più selettiva nelle cellule del glioblastoma multiforme. Fra i molti composti esaminati ne scelsero uno (l'anione borano poliedrico) che si concentrava maggiormente nelle cellule tumorali che in quelle sane circostanti; tuttavia anche questo composto diede risultati deludenti. Tutti i pazienti così trattati presentarono danni di varia gravità ai tessuti cerebrali sani, e specificamente alle pareti dei vasi sanguigni. La concentrazione del boro nella massa tumorale era ancora troppo bassa rispetto a quella nei tessuti sani e nel sangue.

La ricerca di composti del boro che avessero la capacità di localizzarsi selettivamente nelle cellule tumorali continuò utilizzando topi portatori di tumori cerebrali trapiantabili. Questi studi permisero di individuare due composti promettenti, contenenti gruppi solfidrile che sono in grado di legarsi alle proteine. Questa loro proprietà permette forse di spiegare perché essi si concentrino nei tumori cerebrali.

Un'altra area di particolare interesse è lo sviluppo di composti che possano rivelarsi utili nel trattamento del melanoma maligno, una forma pigmentata di neoplasia della cute. L'incidenza di questo tumore e il numero di morti a esso attribuibili sono aumentati di 10 volte negli ultimi 50 anni. Le più recenti statistiche della American Cancer Society prevedono ogni anno 27 600 nuovi casi e 6300 morti attribuibili al melanoma negli Stati Uniti.

Di norma i melanomi possono incorporare grandi quantità di fenilalanina, un amminoacido essenziale, nella melanina e nei suoi precursori, oltre che nelle proteine. Pertanto gli sperimentatori hanno rivolto i loro sforzi alla sintesi di derivati della fenilalanina contenenti boro, nella speranza che questi amminoacidi, paragonabili a cavalli di Troia, vengano utilizzati dalle cellule di melanoma come mattoni biochimici al posto della fenilalanina o in aggiunta a essa.

Un composto che si è dimostrato promettente in animali affetti da melanoma è la *p*-borofenilalanina, un analogo della fenilalanina contenente boro che venne sintetizzato nel 1957 da Harold R. Snyder e collaboratori dell'Università dell'Illinois a Urbana-Champaign. Esso si accumula in quantità significativa non solo nei melanomi, ma anche in altri tipi di tumori. Bernard F. Spielvogel e collaboratori della Boron Biologicals di Raleigh (North Carolina) hanno descritto altri analoghi di amminoacidi contenenti boro, che potrebbero anch'essi rivelarsi utili nella BNCT.

Una linea d'azione analoga tende a sfruttare il fatto che le cellule neoplastiche assumono più facilmente i nucleosidi - i mattoni del DNA - di quanto non facciano le cellule sane. (I nucleotidi sono gli esteri fosforici dei nucleosidi.) Un vantaggio fondamentale offerto dalla concentrazione del boro nel nucleo cellulare è che in tal modo gli effetti distruttivi delle particelle alfa vengono notevolmente incrementati.

Raymond F. Schinazi, ora alla Emory University, ha sintetizzato il primo nucleoside contenente boro in collaborazione con William H. Prusoff. Altri ricercatori negli Stati Uniti e in Giappone stanno mettendo a punto composti correlati che potrebbero concentrarsi maggiormente nel DNA e negli acidi nucleici. Dal momento che la divisione cellulare procede più velocemente nei tumori che nelle cellule sane circostanti, si spera che le cellule neoplastiche incorporino maggiori quantità di nucleosidi contenenti boro e che quindi possano diventare più suscettibili alla reazione di cattura neutronica.

Altre due classi di composti semplici che potrebbero rivelarsi utili come agenti di cattura neutronica nella BNCT sono le porfirine e le promazine. Sebbene siano ancora in fase di sviluppo, diversi di questi composti sono già stati sintetizzati e se ne è dimostrata l'efficacia in colture cellulari e in animali colpiti da tumori.

Un'altra tecnica di introduzione del boro 10 si serve di anticorpi come vettori guidati. Gli anticorpi sono proteine che hanno la potenzialità di riconoscere strutture caratteristiche, dette antigeni, presenti sulla superficie delle cellule tumorali. Vent'anni fa uno di noi (Soloway) e colleghi del Massachusetts General Hospital riuscirono a legare composti contenenti boro a molecole proteiche. All'epoca, però, gli unici anticorpi disponibili non erano abbastanza specifici per introdurre in modo affidabile il boro nelle cellule neoplastiche.

Una quindicina di anni fa, Cesar Milstein e Heinz Kohler dell'Università di Cambridge scoprirono un metodo per ottenere anticorpi di elevata specificità, che vengono chiamati monoclonali. Essi ibridarono cellule normali che producevano anticorpi con cellule di melanoma maligno a rapida divisione; la progenie, composta dai cosiddetti ibridomi, produceva anticorpi monoclonali identici l'uno all'altro.

I ricercatori compresero ben presto che gli anticorpi monoclonali potevano essere utilizzati per introdurre numerosi agenti antitumorali nelle cellule neoplastiche; questa proprietà poteva rivelarsi preziosa anche per far giungere il boro alle cellule tumorali bersaglio. Negli ultimi otto anni noi e altri abbiamo sviluppato metodi per legare composti del boro a molecole di anticorpo. Legando il boro a un polimero dell'amminoacido lisina, per esempio, siamo riusciti a unire più di 1000 atomi di boro a una molecola di anticorpo. Per ragioni ancora poco chiare, le strutture risultanti non sono riuscite a trovare il loro bersaglio antigenico quando sono state inoculate in topi affetti da tumore. Attualmente sono in corso di studio altri metodi che alterano in maniera meno rilevante la molecola di anticorpo.

Di per sé gli anticorpi non sono probabilmente in grado di fornire la quantità necessaria di atomi di boro. Con tutta probabilità verranno usati in combinazione con uno dei composti già noti del boro che si concentrano nelle cellule tumorali. Come i trattamenti chemioterapici efficaci contro il cancro si basano sull'associazione di farmaci che attaccano in vari modi le cellule neoplastiche, così, per avere successo, la BNCT dovrà dipendere da metodi diversi per introdurre la quantità necessaria di boro-10 nelle cellule tumorali.

Questa ricerca ha dato tuttavia un risultato immediatamente sfruttabile. Insieme con Moise Bendayan dell'Università di Montreal, in Canada, abbiamo scoperto che il boro può essere impiegato insieme con uno speciale tipo di microscopio elettronico, in grado di rilevare singoli atomi. In questa applicazione il boro si dimostra una sonda ideale per identificare le strutture e i componenti chimici delle cellule.

Il successo della BNCT dipende indubbiamente da diversi altri fattori, tra cui la posizione e la profondità del tumore, la concentrazione relativa di boro 10 nel tumore e nei tessuti sani e il numero e l'energia dei neutroni che raggiungono il sito bersaglio. I calcoli hanno dimostrato che sono necessari un miliardo di atomi di boro per cellula e un flusso di neutroni termici di 10^{12}-10^{13} neutroni al centimetro quadrato. Il motivo per cui occorre un numero così elevato di atomi di boro è l'esigenza di far sì che la dose di radiazione ricevuta sotto forma di particelle alfa superi la radiazione di fondo prodotta dalle reazioni di cattura neutronica dell'azoto e dell'idrogeno.

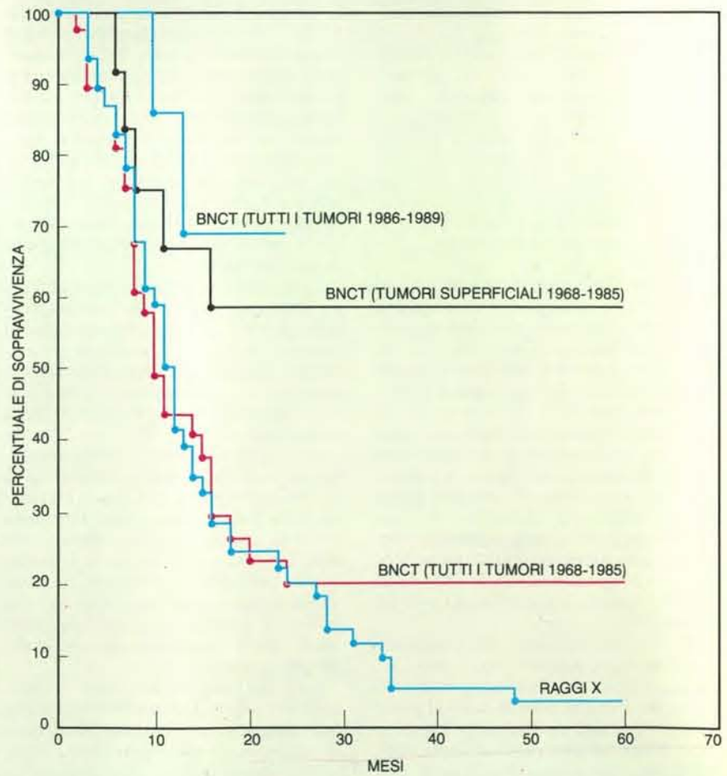
La necessità di distruggere un numero di cellule neoplastiche sufficiente a impedire la ricrescita del tumore spiega perché occorra un flusso di neutroni così intenso. Con la BNCT si dovrebbe portare la dose di radiazione ricevuta dal tumore a livelli più alti di quelli raggiungibili con la radioterapia convenzionale, pur mantenendo la dose ricevuta dai tessuti sani inferiore al livello di tolleranza.

I neutroni sono generati nei reattori nucleari in seguito a reazioni di fissione e la loro energia iniziale è all'incirca di un milione di elettronvolt. È possibile rallentarli fino alle energie tipiche dei neutroni termici tramite moderatori quali l'acqua pesante o l'acqua comune. Tuttavia i neutroni termici pongono alcuni problemi. Essi vengono rapidamente attenuati mentre diffondono nei tessuti ed è quindi difficile farne giungere un numero sufficiente alla profondità richiesta senza irradiare in maniera eccessiva i tessuti superficiali.

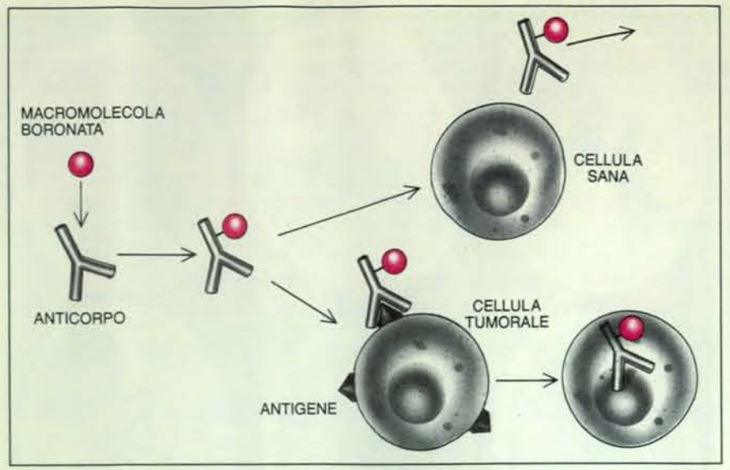
Una possibile alternativa è di operare con fasci di neutroni epitermici, aventi energie fino a qualche migliaio di elettronvolt. Questi neutroni hanno maggiore capacità di penetrazione e quindi non danneggiano quanto i neutroni termici i tessuti superficiali sani. Oltre a ciò i neutroni epitermici perdono energia nell'attraversare i tessuti e risultano

NUCLIDE		BARN
BORO 10	^{10}B	3838,00
ELEMENTO		
OSSIGENO	O	0,0002
CARBONIO	C	0,0037
MAGNESIO	Mg	0,069
FOSFORO	P	0,19
IDROGENO	H	0,332
CALCIO	Ca	0,44
ZOLFO	S	0,52
SODIO	Na	0,536
AZOTO	N	1,75
POTASSIO	K	2,07
FERRO	Fe	2,62
CLORO	Cl	33,8

Come il boro anche molti altri elementi presenti nell'organismo catturano i neutroni con emissione di radiazioni. Ma la sezione d'urto di cattura di questi elementi, ossia la loro probabilità di catturare neutroni (misurata in barn, pari a 10^{-24} centimetri quadrati), è bassa rispetto a quella del boro.



Una terapia, ancora in fase sperimentale, dei tumori cerebrali è stata messa a punto in Giappone da Hiroshi Hatanaka della Teikyo University. Un tumore (*fila superiore di tomografie*) trattato con intervento chirurgico e BNCT non ha presentato recidive (*fila inferiore di tomografie*). Il diagramma in basso riassume i risultati ottenuti nel periodo compreso tra il 1968 e il 1989 su un centinaio di pazienti. La combinazione di intervento chirurgico e BNCT ha consentito una sopravvivenza più lunga rispetto al trattamento con raggi X.



Gli anticorpi monoclonali, grazie alla elevata specificità che li caratterizza, sono in grado di trasportare selettivamente il boro nelle cellule cancerose agendo come vettori guidati. Il medesimo scopo può essere raggiunto anche mediante farmaci o composti naturali.

«termalizzati» alle profondità alle quali i tumori tendono a formarsi.

Apparecchiature per produrre fasci di neutroni epitermici adatti alle applicazioni cliniche sono state installate o sono in corso di installazione al Brookhaven National Laboratory e al MIT. Oltre a ciò, i reattori dell'Università del Missouri a Columbia, del Georgia Institute of Technology e dell'Idaho National Engineering Laboratory sono in grado di produrre fasci epitermici efficaci.

Se le sperimentazioni cliniche ad ampio raggio dimostreranno la validità della BNCT, lo sviluppo di fonti alternative di neutroni assumerà notevole importanza a causa dell'alto costo e del numero limitato di reattori nucleari, oltre che dell'opposizione alla costruzione di nuovi reattori nelle aree metropolitane. Esistono tuttavia metodi alternativi per generare neutroni, fra i quali i radioisotopi, gli acceleratori di protoni e i processi di spallazione del nucleo. (Nei processi di spallazione vengono emessi neutroni e altre particelle leggere come prodotti di reazioni nucleari in seguito al bombardamento con protoni di elementi pesanti.)

Fra i vari radioisotopi che producono neutroni, il più adatto è il californio 252, che dà origine per fissione spontanea a neutroni di energia simile a quelli generati nei reattori nucleari. Per ottenere fasci di intensità adeguata all'applicazione terapeutica saranno necessari diversi grammi dell'isotopo. Avendo un tempo di dimezzamento di 2,7 anni, il californio 252 potrebbe costituire una fonte di neutroni semplice, sicura e affidabile.

Gli acceleratori di protoni a bassa energia potrebbero generare neutroni epitermici mediante irradiazione del litio. James W. Blue e suo figlio Thomas E. Blue, che lavorano rispettivamente al Lewis Research Center della National Aeronautics and Space Administration a Cleveland e alla Ohio State University, hanno calcolato che sarebbero necessarie correnti del fascio di protoni vicine a due milliampere. Correnti del genere sono tecnicamente ottenibili, ma finora non sono mai state raggiunte. Sarà necessario moderare il fascio con acqua pesante per ottenere un numero adeguato di neutroni epitermici.

La spallazione di elementi pesanti come rame, piombo e uranio produce neutroni di varie energie; le particelle di energia più elevata possono essere moderate a livelli epitermici. Misurazioni effettuate sulla spallazione del rame, tuttavia, indicano che il processo richiede diverse ore.

Noi e altri ricercatori siamo stati incoraggiati a portare avanti i nostri tentativi dai risultati ottenuti da Hiroshi Hatanaka della Teikyo University. Hatanaka iniziò la sperimentazione clinica della BNCT in Giappone 22 anni fa. Durante questo periodo ha sottoposto a trattamento più di 100 pazienti colpiti da varie forme di tumori cerebrali, compreso quello che si mostra più resistente alla terapia, il glioblastoma.

La prima fase del trattamento di Hatanaka consiste nella rimozione chirurgica della maggiore quantità possibile di massa tumorale. Dopo l'operazione, al paziente viene inoculato attraverso un'arteria l'anione borano poliedrico di formula $Na_2B_{12}H_{11}SH$. Dopo che il boro ha raggiunto le cellule neoplastiche, il sito tumorale viene irradiato con un fascio focalizzato di neutroni termici.

Purtroppo l'efficacia di questo trattamento è ridotta da svariate limitazioni. Una di esse è il fatto che il reattore nucleare usato come fonte di neutroni ha una potenza così bassa che l'irradiazione può durare anche tre-cinque ore. Oltre a ciò, la profondità di penetrazione efficace è meno di sei centimetri, cosicché il trattamento si applica solo a tumori superficiali.

Nondimeno, i risultati di Hatanaka sono degni di nota per diverse ragioni. In primo luogo, solo uno dei 100 pazienti ha presentato danni da radiazione ai tessuti cerebrali sani, e quel paziente aveva ricevuto una dose di neutroni eccezionalmente elevata. Inoltre, in 38 pazienti con glioblastoma multiforme e in 12 con tumori più superficiali, la sopravvivenza media è stata significativamente maggiore che nei pazienti trattati con la radioterapia convenzionale. Diversi pazienti sono sopravvissuti per anni, mentre la sopravvivenza media dei soggetti affetti da questo tipo di tumore è inferiore a un anno.

La sperimentazione della BNCT su gliomi trapiantati nel cervello di ratti conferma i dati clinici di Hatanaka. Forse ancora più significativo è il fatto che Darrel D. Joel e colleghi del Brookhaven National Laboratory hanno studiato una forma dimera dell'anione borano poliedrico $Na_2B_{12}H_{11}SH$ la cui affinità per i tumori sembra maggiore di quella del monomero. Essi hanno osservato che il dimero è efficace nel trattamento di un tumore cerebrale del ratto. Diversi gruppi di ricerca negli Stati Uniti e in Europa hanno avviato le prime fasi di un programma che potrebbe svilupparsi in una sperimentazione clinica controllata della BNCT per il glioblastoma multiforme.

Diversi recenti sviluppi hanno notevolmente incrementato le potenzialità della BNCT. Sono oggi disponibili nuovi composti del boro che presentano una rilevante affinità per i tessuti cancerosi e una chiara efficacia nel trattamento di tumori negli animali, e sono stati messi a punto metodi per ottenere fasci di neutroni epitermici che consentano il trattamento di tumori profondi.

Yutaka Mishima e colleghi della Kobe University School of Medicine hanno compiuto i maggiori progressi nell'applicazione della BNCT al melanoma. Utilizzando un particolare ceppo di cavie che sviluppano spontaneamente melanomi, i ricercatori hanno inoculato un composto del boro, la ^{10}B-BPA (^{10}B-borofenilalanina) intorno a un melanoma cutaneo. L'inoculazione è stata seguita da un'unica esposizione a neutroni termici e, nel giro di due mesi, i melanomi hanno iniziato a ridursi fino a sparire completamente.

Questi risultati hanno condotto a una sperimentazione clinica ora in corso. Fino alla primavera 1990 erano stati trattati otto pazienti con melanoma cutaneo che per varie ragioni non potevano essere sottoposti a intervento chirurgico. Mishima ha riferito che in due pazienti si è avuta la guarigione dai melanomi primari, mentre gli altri soggetti trattati sono seguiti costantemente per individuare eventuali forme recidive.

Più difficoltosa è l'applicazione della BNCT ai melanomi sottocutanei. Jeffrey A. Coderre e colleghi del Brookhaven National Laboratory hanno trattato con BNCT topi e ratti portatori di tumori trapiantabili utilizzando la ^{10}B-BPA. Si è osservato che con la dose minima di radiazione la crescita tumorale è rallentata, mentre con la dose massima i tumori cessano di crescere o addirittura regrediscono spontaneamente.

L'applicazione della BNCT potrebbe essere estesa ad altri tumori sui quali le attuali terapie hanno scarso successo. In effetti, se si riusciranno a superare tutti gli ostacoli, la BNCT potrebbe divenire disponibile presso centri specializzati dotati di acceleratori nucleari o presso ospedali che abbiano in dotazione fonti alternative di neutroni. Nel caso di tumori letali come i glioblastomi, per i quali oggi non esiste alcuna terapia efficace, la BNCT potrebbe divenire in futuro il trattamento di elezione sia da sola, sia come terapia coadiuvante, in associazione con l'intervento chirurgico e la chemioterapia oppure la radioterapia convenzionale.

Solo un miglioramento della tecnica come risultato delle ricerche attualmente in corso nei laboratori di tutto il mondo permetterà alla BNCT di trovare un posto nella terapia dei tumori. Comunque, quali che siano i risultati di questi studi, è probabile che i sistemi terapeutici a due componenti diano un contributo importante al trattamento del cancro nel prossimo secolo.

BIBLIOGRAFIA

HATANAKA HIROSHI (a cura), *Boron-Neutron Capture Therapy for Tumors*, Nishimura Co., Ltd., 1986.

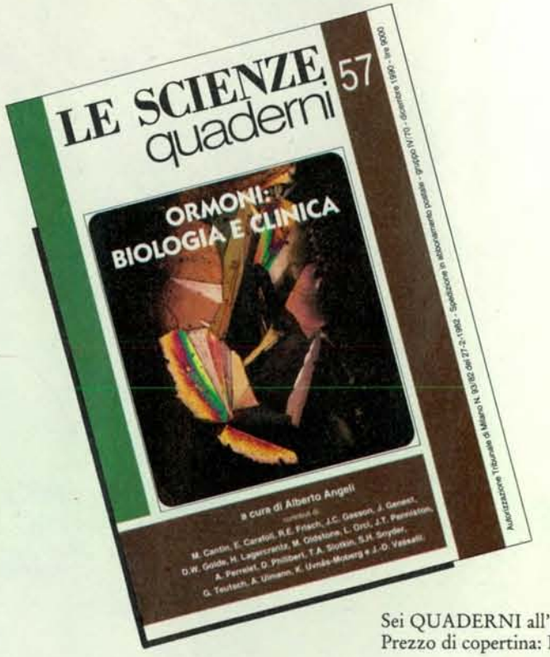
GABEL DETLEF (a cura), *Third International Symposium on Neutron Capture Therapy* in «Strahlentherapie und Onkologie», 165, n. 2/3, febbraio/marzo 1989.

FAIRCHILD RALPH G., BOND V. P. e WOODHEAD A. D. (a cura), *Clinical Aspects of Neutron Capture Therapy*, Plenum Press, 1989.

MISHIMA YUTAKA (a cura), *The Second Japan-Australia International Workshop on Thermal Neutron Capture Therapy for Malignant Melanoma*, Kobe, 23-26 settembre 1987, in «Pigment Cell Research», 2, n. 4, pp. 223-386, 1989.

BARTH ROLF F., SOLOWAY ALBERT H. e FAIRCHILD RALPH G., *Boron Neutron Capture Therapy of Cancer* in «Cancer Research», 50, n. 4, 15 febbraio 1990.

Alla scoperta del linguaggio interno del nostro corpo è dedicato il nuovo quaderno di «Le Scienze» a cura di Alberto Angeli
ORMONI: BIOLOGIA E CLINICA
in edicola e in libreria da dicembre.



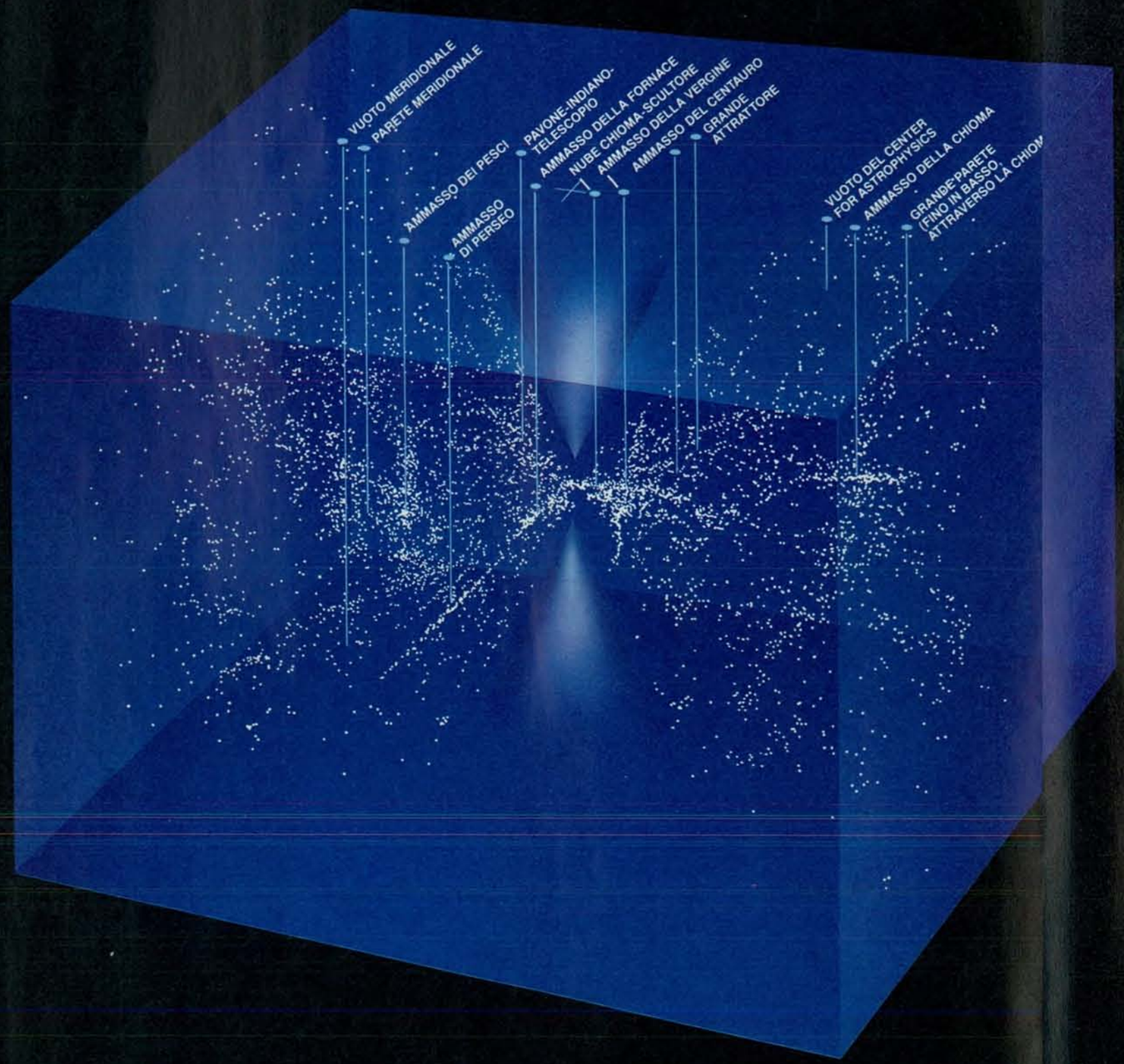
Sei QUADERNI all'anno.
Prezzo di copertina: L. 9000.

In questo numero:

Le basi molecolari della comunicazione fra cellule di S.H. Snyder
Il calcio come messaggero biologico di E. Carafoli e J.T. Penniston
Peso corporeo e fertilità di R.E. Frisch
Lo stress della nascita di H. Lagercrantz e T.A. Slotkin
Ritmi endocrini nell'uomo di A. Angeli
Gli ormoni dell'apparato gastrointestinale di K. Uvnäs-Moberg
Il cuore come ghiandola endocrina di M. Cantin e J. Genest
Le emopoietine di D.W. Golde e J.C. Gasson
La fabbrica dell'insulina di L. Orci, J.-D. Vassalli e A. Perrelet
Virus che alterano la funzione della cellula di M. Oldstone
La pillola R.U. 486: un antiormone di A. Ulmann, G. Teutsch e D. Philibert

COSMOLOGI A CONVEGNO

di John Horgan



Questa mappa tridimensionale di 6500 galassie mostra ammassi di galassie, vuoti e altre strutture di grande scala dell'universo locale. La posizione approssimativa delle strutture più importanti è stata proiettata sulla faccia superiore del parallelepipedo. Le regioni coniche ombreggiate non si vedono a causa della Via Lattea, che si trova al centro (al di sotto della croce). L'immagine, realizzata per «Scientific American» da Changbom Park della Princeton University si basa su dati provenienti da vari rilevamenti degli spostamenti verso il rosso: quello del cielo australe (Southern Sky), quello di Perseo e dei Pesci e quelli effettuati dallo Harvard-Smithsonian Center for Astrophysics. La mappa rappresenta una regione di 650 milioni di anni luce di lato e dell'altezza di 450 milioni di anni luce; se contenesse gli oggetti più lontani mai osservati sarebbe all'incirca 50 volte più grande.

La teoria cosmologica maggiormente accreditata, il modello a materia oscura fredda, è stata oggetto di vivaci contestazioni a un recente convegno tenutosi in Svezia.

Chi ha buttato sabbia sullo schermo? In realtà si tratta della proiezione del lucido che rappresenta l'universo secondo Carlos S. Frenk, astronomo dell'Università di Durham, nel Regno Unito. Più esattamente si tratta di una simulazione al calcolatore che mostra come la materia si sia condensata sotto l'effetto della gravità dopo il big bang, l'esplosione cosmica che miliardi di anni fa ha dato origine all'universo.

Ogni granello dell'immagine è tutta una galassia, già di per sé un piccolo universo con i suoi miliardi di stelle. Frenk, quest'uomo corpulento dall'accento messicano, sostiene che la sua simulazione assomiglia moltissimo alle carte vere della distribuzione delle galassie: «Il nostro modello - dichiara con una ferma espressione di sfida - riesce a riprodurre qualunque struttura di grande scala osservabile.»

È comprensibile che Frenk appaia un po' teso: il pubblico che lo ascolta è costituito da oltre 30 tra i maggiori fisici e astronomi del mondo, convenuti in un austero albergo nel nord della Svezia, terra montuosa e subartica, per un raduno organizzato dalla Fondazione Nobel volto a fare il punto sulla nascita e le prime fasi dell'evoluzione dell'universo. Per sei giorni a metà di giugno questi scienziati non faranno quasi altro che parlare di cosmologia, a tavola di fronte alle grigliate di renna, durante le passeggiate su un'altura dei dintorni (battezzata «il massimo locale» da uno dei fisici) o durante le sessioni ufficiali programmate per ciascuna giornata.

Frenk sa bene che alcuni di questi studiosi del cosmo stanno diventando sempre più scettici sulla teoria che sta alla base della sua simulazione, quel modello a materia oscura fredda basato sull'idea piuttosto inquietante che almeno il 90 per cento della massa complessiva dell'universo sia costituito da qualche forma di materia invisibile. Tutte le stelle e le galassie luminose che si osservano al telescopio, in altre parole, non sarebbero che la spuma delle onde su un oceano buio. Il modello è la descrizione più diffusa dell'evoluzione dell'universo dopo il big bang, tanto che uno dei fisici presenti lo chiama «canone».

Le ultime osservazioni però incalzano il modello con una manovra a tenaglia. Da una parte, le misurazioni sempre più precise del fondo cosmico a microonde - un bagno di microonde freddo (2,7 gradi sopra lo zero assoluto) che si pensa sia il debole bagliore residuo del big bang - non riescono ancora a rivelare alcuna variazione di intensità a livello regionale.

ISTANTANEE DA UN CONVEGNO

Un giornale locale, colpito dalla straordinarietà dell'avvenimento, lo ha descritto come un incontro tra i «più grandi genî del mondo». Dall'11 al 16 giugno, più di 30 scienziati si sono riuniti in un austero albergo del nord della Svezia per un convegno sulla nascita e le prime fasi dell'evoluzione dell'universo. Ma anche i genî non vivono di sole elucubrazioni; le giornate del convegno sono quindi state allietate da simpatiche riunioni conviviali e da brevi escursioni in zona. Una sera, per esempio, sono andati tutti insieme in una fattoria ad assaggiare la renna, l'alce e l'orso e ad assistere a canti e danze popolari messi in scena dalla gente del luogo, mentre un'altra volta in elicottero hanno raggiunto un lago di montagna per un brindisi a base di punch al ribes selvatico. Inoltre, vari gruppetti, costituiti senza tener conto delle compatibilità teoriche, hanno esplorato le montagne circostanti.

Violinisti svedesi, Stephen Hawking e la sua assistente

Jim

Peebles (*a sinistra*) e Carlos Frenk

Rocky Kolb (*a sinistra*) e Alan Guth

Dave Schramm (*a sinistra*) con alcuni ospiti svedesi

Un gruppo di cosmologi

John Preskill, Frank Wilczek

e Gerard 't Hooft

Andrei Linde

Neil Turok

Questo costringe gli autori del modello a ipotizzare che all'inizio l'universo fosse estremamente omogeneo, con la materia distribuita uniformemente in tutto lo spazio. Contemporaneamente, d'altro canto, le carte dell'universo rivelano addensamenti di galassie sempre più grandi circondati da vuoti di dimensioni pure crescenti: ma se l'universo era così omogeneo all'inizio, come mai ora è tutto a grumi? È lodevole lo sforzo che i sostenitori del modello a materia oscura fredda stanno facendo per rispondere a questa domanda.

Un critico del modello è P. James E. Peebles della Princeton University, un personaggio importante del panorama cosmologico fin dalla metà degli anni sessanta. Alto e magrissimo, osserva le simulazioni di Frenk stando appollaiato come un falco al centro della prima fila, e appena questi fa per tornare al suo posto balza in piedi e sfodera il suo lucido. Sullo schermo appare una fascia densa di galassie con intorno solo pochi punti sparsi. Non si tratta di una simulazione, avverte Peebles, ma della realtà, della «metropoli» di galassie in cui si trova la Via Lattea, la nostra galassia. Parlando evidenzia con dei cerchi i grandi spazi vuoti sopra e sotto la fascia centrale: «Dove sono questi vuoti nel tuo modello, Carlos?» chiede, puntando su Frenk gli occhiali cerchiati di metallo. «E queste?» ribatte Frenk, tornato alla lavagna luminosa per indicare le galassie sparse. «Io preferisco concentrarmi su questo» asserisce Peebles e batte l'indice sul vuoto più grande. «Ci hai sempre insegnato a essere quantitativi» esclama Frenk, e Peebles di rimando, con il tono di chi rimprovera uno studente poco pronto: «Faresti bene a essere anche ragionevole!».

Un testa a testa fra cosmologi, proprio come speravano i fisici del Politecnico Chalmers quando hanno organizzato il convegno per la Fondazione Nobel. La compagnia riunita è variegata e polemica, e le conferenze sfuggono di rado alle interruzioni. Quando Lennox L. Cowie dell'Università di Hawaii cerca di interpretare le sue ultime immagini di galassie nello spazio profondo il pubblico lo bombarda di domande: «Quando parli di "formazione delle galassie" vuoi dire "formazione delle stelle", vero?», «Perché non c'è l'elio in quella tabella?», o ancora «Ti spiace toglierti da davanti allo schermo?». Cowie, un grosso scozzese con un faccione a luna piena, ricorre all'autoironia: «Tutte le idee intelligenti che presento qui sono probabilmente rubate a Jim Peebles - dichiara, e nel suo accento verace si avverte un tremito - e tutte quelle completamente idiote sono invece mie.»

Forse l'unica idea sulla quale sembra che nessuno nutra dubbi è la premessa fondamentale della cosmologia classica: la teoria del big bang, che poggia su tre solidi pilastri sperimentali. Il primo è la scoperta, compiuta dall'astronomo statunitense Edwin P. Hubble 60 anni fa, che la luce emessa dalle galassie è spostata verso il rosso, cioè verso le frequenze più basse dello spettro, in proporzione alla loro distanza da noi. Proprio come l'abbassamento di tono del fischio di un treno rivela che questo si sta allontanando da noi, lo «spostamento verso il rosso» della luce dimostra che le galassie si stanno allontanando e il fatto che la velocità di allontanamento sia proporzionale alla distanza prova che l'universo stesso si sta espandendo. (La correlazione tra velocità di allontanamento e distanza reciproca vale anche per i punti di un elastico: tirare per credere.) Estrapolando a ritroso la velocità di espansione attuale, gli astronomi hanno dedotto che il big bang si sarebbe verificato 13 miliardi di anni fa, più o meno ben sette miliardi di anni. L'incertezza delle stime dipende dall'imprecisione delle valutazioni della distanza delle galassie, della massa dell'universo (che modifica la velocità di espansione) e di altri parametri ancora.

La seconda prova sperimentale del big bang è la radiazione di fondo a microonde, scoperta nel 1965 da Arno A. Penzias e Robert W. Wilson dei Bell Telephone Laboratories. La terza, infine, è costituita dalla composizione osservata dell'universo: l'idrogeno rappresenta circa tre quarti del totale in massa e l'elio un quarto, con tutti gli altri elementi presenti in tracce. Vari fisici hanno calcolato che un big bang caldo avrebbe prodotto proprio questo miscuglio di elementi tramite il normale processo di fusione nucleare, ossia per nucleosintesi.

«Il big bang sta benissimo» proclama David N. Schramm dell'Università di Chicago, che ha dato un contributo importantissimo all'affinamento dei calcoli sulla nucleosintesi, e viene proprio voglia di dargli ragione. Secondo i colleghi «Schrammbo», così soprannominato per la sua abilità di pilota, di alpinista e anche per il suo passato di campione di lotta grecoromana, è, a detta dei colleghi, il cosmologo di «maggior peso» a livello mondiale, secondo solo a Geoffrey R. Burbidge dell'Università della California a San Diego. Schramm è anche un veemente entusiasta del suo campo. Altri di fronte alle disgrazie della materia oscura fredda e di tutte le altre teorie sulla formazione delle galassie si perdono d'animo e paventano la crisi, ma lui parla di «età dell'oro» sottolineando che «la struttura di base c'è; ora bisogna solo riempire i buchi».

Ma in questi buchi si annidano problemi importanti: perché la radiazione di fondo a microonde è così uniforme? Quanta materia è presente nell'universo, e di che cosa è fatta? Quando si sono formate le galassie, e come? L'universo continuerà a espandersi per sempre o finirà per collassare ricadendo su se stesso?

I fisici teorici hanno affrontato queste domande in vari modi. Alcuni hanno elaborato l'ipotesi che le galassie e le strutture più grandi siano nate da certi difetti nella trama dello spazio, le corde (o stringhe) cosmiche, o da oggetti analoghi, le «intersezioni», o da esplosioni potentissime verificatesi ai primordi della storia dell'universo. Altri scienziati ritengono che la chiave di tanti paradossi cosmici stia in un'ombra di energia posseduta dallo spazio vuoto, la costante cosmologica. I cultori della «cosmologia quantistica», dal canto loro, cercano le risposte fondamentali nel primo istante della creazione, quando tutte le forze della natura erano unificate.

Frenk, per esempio, pensa che quella della materia oscura fredda sia ancora la soluzione più promettente. Incalzato da vicino arriva ad ammettere che con le ultime osservazioni il modello è giunto al punto di rottura e probabilmente avrà bisogno, come minimo, di una revisione, ma continua a sostenere che le teorie alternative finiscono di norma per essere ancora meno compatibili coi dati o ancora troppo poco sviluppate per poter essere confrontate con i risultati sperimentali. Viene in mente quel vecchio detto statunitense secondo il quale chi è in testa alla corsa è quello con tutte le frecce piantate nella schiena.

Giusto o sbagliato che sia, il modello della materia oscura fredda è un cocktail incredibilmente ardito e fantasioso di idee prese dalla fisica delle particelle e dalla relatività generale. Forse la sua componente più cruciale è il concetto di inflazione, tanto che alcuni astronomi parlano di un modello a materia oscura fredda più inflazione.

Guth e le sue teorie

Il primo a formulare il concetto di inflazione, una decina di anni fa, è stato Alan H. Guth, un personaggio dal faccione rotondo con il naso a patata e un taglio di capelli alla Beatles prima maniera. Guth, a giudicare dall'aria, sembra più uno studente di quelli che smanettano sul computer tutta la notte che un professore del Massachusetts Institute of Technology. Al convegno svedese Guth manifesta la tendenza ad assopirsi durante i discorsi altrui (e non si può non notare dato che siede in prima fila). Questo permette a uno dei fisici di sfoggiare un certo orgoglio nel proclamare, poco dopo l'inizio del proprio intervento, di avere «ottemperato al criterio di Guth in forma debole: riuscire a tenere Alan sveglio per almeno cinque minuti». E Guth è quello che, alla battuta, ride più rumorosamente.

Guth è però estremamente acuto nell'escogitare nuove ipotesi, e nel propagandarle. Nel concetto di inflazione s'è imbattuto per caso analizzando certe teorie di grande unificazione (dette anche GUT) della fisica delle particelle secondo le quali a energie elevate (quindi ad alte temperature) l'interazione nucleare forte, quella debole e l'elettromagnetismo, cioè tutte le forze della natura tranne la gravità, si fondono in una sola. Queste temperature sono assolutamente irraggiungibili da qualsiasi esperimento eseguito sulla Terra, ma devono essere state comuni nell'universo qual era 10^{-35} secondi dopo il big bang, un periodo chiamato era delle teorie di grande unificazione, o «era GUT».

L'era delle teorie di grande unificazione finisce quando l'interazione nucleare forte si separa dalle altre forze e l'universo subisce una transizione di fase simile alla trasformazione dell'acqua in ghiaccio. Secondo l'ipotesi di Guth, verso la fine di questo periodo, la transizione potrebbe essere ritardata in certe regioni dell'universo, lasciandole in uno stato di sovraraffreddamento, proprio come l'acqua quando rimane liquida anche al di sotto del punto di congelamento. Secondo la sua proposta queste regioni sovraraffreddate sarebbero entrate in uno stato speciale detto «falso vuoto» nel quale la gravità diventa una forza repulsiva anziché attrattiva, subendo un'espansione breve, ma intensissima: l'in-

flazione. Una di queste regioni, inizialmente più piccola di un protone, ma cresciuta in un istante alle dimensioni di un pompelmo, sarebbe poi diventata il nostro universo.

Secondo questo scenario tutto l'universo osservabile non è che una bollicina in un cosmo molto più grande, e quasi tutta la materia che la costituisce è stata creata virtualmente dal nulla. Proponendola, Guth ha sottolineato che questa teoria, oltre a far vacillare la mente, riesce anche a risolvere alcuni problemi cosmologici fastidiosi. Uno di essi, chiamato talvolta anche «problema dell'omogeneità», deriva dai calcoli basati sul modello del big bang pre-inflazione, secondo i quali la palla di fuoco primordiale si sarebbe espansa troppo in fretta per permettere alle sue regioni di scambiarsi radiazione e quindi raggiungere l'equilibrio termico. Com'è possibile allora, chiedevano alcuni cosmologi, che la luminescenza residua del big bang, la radiazione di fondo a microonde, abbia la stessa temperatura in tutte le direzioni? Secondo Guth l'inflazione risolve il problema: la regione da cui ha preso forma tutto l'universo osservabile era in origine talmente piccola da essere già in equilibrio fin dall'inizio.

L'inflazione ha fatto giustizia anche del «problema della curvatura nulla», legato alla teoria della relatività generale di Einstein, secondo la quale la materia, per il tramite della gravità, fa curvare lo spazio. Da questo effetto dipende il destino ultimo di un universo in espansione. Se infatti la quantità di materia per unità di volume, cioè la densità di massa, dell'universo supera un certo livello critico, lo spazio si curva su se stesso in modo tale che le linee parallele convergono. La conseguenza più rilevante è che la gravità prima o poi arresterà l'espansione e farà ricontrarre l'universo in un «*big crunch*», cioè un «catafascio generale», volendo rendere in italiano questa definizione di cui Schramm reclama la paternità. Un universo di questo tipo si dice chiuso. Se invece la densità di massa è inferiore a quella critica, le linee parallele divergono e l'universo, detto aperto, si espanderà indefinitamente. Infine, se la densità ha esattamente il valore critico, l'universo continuerà a espandersi, ma a una velocità sempre decrescente. Quest'ultimo tipo di universo sarebbe «piatto», non avendo lo spazio alcuna curvatura; qui le linee parallele si comportano proprio come diceva Euclide e come sembra avvenire nel mondo che conosciamo.

Prima della teoria inflazionaria diversi cosmologi avevano osservato che l'apparente curvatura nulla dell'universo, e perfino la sua stessa esistenza, rappresentano una coincidenza straordinaria: se l'universo avesse contenuto appena un pizzico di materia in più di quella che c'è, sarebbe collassato moltissimo tempo fa; se ne avesse contenuta poco meno, si sarebbe volatilizzato troppo in fretta per dar tempo alle stelle e alle galassie di formarsi. Come ha detto Lawrence M. Krauss, della Yale University, la probabilità di azzeccare la densità esatta per l'universo è pari alla probabilità di indovinare quanti atomi ci sono nel Sole: come abbiamo fatto ad avere tanta fortuna?

Ancora una volta la teoria inflazionaria ci offre una risposta: Guth ha sostenuto che l'inflazione avrebbe spianato la superficie di quella parte dell'universo che riusciamo a osservare, proprio come la superficie di un pallone gonfiato a dimensioni 1000 volte maggiori di quelle normali sembrerebbe piatta guardandola da vicino. In un universo così piatto la densità verrebbe ad assumere automaticamente il valore critico.

Misteri come quelli dell'omogeneità e della curvatura nulla hanno indotto alcuni cosmologi a invocare il principio antropico: l'universo è fatto proprio così perché se fosse diverso non ci saremmo noi a osservarlo. Ma quasi tutti gli scienziati considerano questo principio tautologico e ascientifico: «Invocando il principio antropico - dice Schramm - si afferma solo di non saper risolvere il problema.» Non stupisce quindi che numerosi fisici teorici abbiano abbracciato la teoria inflazionaria salutandola come lo sviluppo più importante dopo l'originale formulazione della teoria del big bang. Essa, infatti, non solo risolve i problemi dell'omogeneità e della curvatura nulla, ma formula previsioni specifiche sulla quantità di materia presente nell'universo e sul suo destino ultimo.

La teoria inflazionaria spiega persino la formazione delle galassie. Secondo la fisica quantistica, infatti, a livello subatomico l'intensità di qualunque campo di energia fluttua sempre, proprio come le piccole onde che increspano perennemente la superficie di un lago anche nella calma più piatta. Dopo l'inflazione i rilievi inizialmente creati nell'universo da questo effetto sarebbero cresciuti abbastanza per fungere da «semi» per la nascita di stelle e galassie; la gravità poi avrebbe fatto il resto. «Molti pensano che l'inflazione sia un concetto troppo bello per non essere giusto in una forma o nell'altra», dice Schramm.

L'inflazione ha risolto alcuni gravi problemi, ma ne ha sollevato un altro: dov'è tutta la materia? La teoria di Guth prevede senza possibilità di equivoco che la densità della materia nell'universo sia esattamente quella critica, ma tutte le misurazioni compiute hanno dato risultati molto inferiori.

Le indagini sulla materia visibile, cioè luminosa, hanno fornito per la densità di massa un valore pari all'1 per cento appena del valore critico. Lo studio dei moti stellari all'interno delle singole galassie ha fatto salire le stime, perché le osservazioni indicano che le galassie devono contenere circa 10 volte più massa di quanto non sembri, altrimenti andrebbero a pezzi; l'analisi del moto delle galassie all'interno degli ammassi ha portato alla stessa conclusione.

Questi risultati, ottenuti per la prima volta vari decenni fa, hanno dato vita alla ricerca della «materia mancante», ma la revisione delle stime per la massa delle galassie e degli ammassi ha fatto salire quella della densità complessiva dell'universo solo al 10 per cento circa del valore critico. Per di più, un assunto fondamentale dei calcoli sulla nucleosintesi primordiale (quelli che spiegano così bene le abbondanze relative degli elementi leggeri nell'universo) è che i barioni, la famiglia di particelle che comprende i protoni e i neutroni e rappresenta la quasi totalità della materia di cui sono fatte le stelle, i pianeti e le persone, non possano assolutamente superare il 10 per cento della densità critica, altrimenti il big bang avrebbe prodotto più elio.

Questa affermazione ha conseguenze profonde. All'inizio gli astronomi sospettavano che la materia oscura fosse costituita esclusivamente da materia barionica, normale, in forme difficili da individuare: comete, pianeti, stelle molto poco luminose chiamate nane brune o marroni, buchi neri e persino intere galassie troppo deboli per essere osservate. Ma, se i calcoli dell'inflazione e quelli della nucleosintesi sono corretti, almeno il 90 per cento della materia dell'universo deve essere costituito da particelle non barioniche, e per di più queste particelle devono avere interazioni molto deboli con la materia comune, altrimenti non sarebbero così difficili da rilevare.

All'inizio il candidato favorito per il ruolo di materia oscura era una particella non barionica «calda», cioè veloce: il neutrino. I neutrini vengono prodotti in grandi quantità in varie reazioni nucleari, ma presentano interazioni molto deboli, tanto da poter attraversare tutta la Terra senza risentirne affatto. Il fatto importante è che di sicuro ne è stata dimostrata l'esistenza: la prima osservazione di uno dei tre tipi di neutrini che si conoscono risale al 1956.

I modelli a dominante neutrinica, altrimenti detti a «materia oscura calda», pongono però due problemi. Innanzitutto anni e anni di esperimenti non sono riusciti a dimostrare che i neutrini abbiano una massa, meno che mai, quindi, a valutare se questa sia sufficiente a frenare l'espansione dell'universo. Ultimamente i primi risultati provvisori ottenuti da un nuovo rivelatore di neutrini ad alta sensibilità, il Soviet-American Gallium Experiment (SAGE), hanno rianimato le speranze di eliminare il problema nutrite dai sostenitori della materia oscura calda: i dati del SAGE inducono a ritenere che il Sole emetta molti meno neutrini di quanti ne prevede il modello standard della fisica delle particelle e ammettere che un tipo di neutrini sia dotato di massa è uno dei modi possibili di spiegare questa discrepanza. Resta però da risolvere un altro problema: secondo le simulazioni al calcolatore i neutrini, così veloci, ci avrebbero messo troppo tempo per assestarsi in galassie.

Le carenze dei modelli a materia oscura calda hanno portato i teorici a concentrarsi su un tipo di materia oscura più lenta, fredda, che dovrebbe addensarsi più rapidamente. Il primo a scrivere sull'inflazione con materia oscura fredda fu, nel 1982, Peebles, ma in seguito divenne critico perché - come sostiene - il modello ha «troppi problemi». Frenk dice che Peebles si diverte a proporre teorie per poi criticarle quando hanno successo. «È una leggenda - replica Peebles - ma come tutte le leggende potrebbe nascondere un fondo di verità.»

La banda dei quattro

Intorno alla metà degli anni ottanta Frenk, Marc Davis dell'Università della California a Berkeley, George Efstathiou dell'Università di Oxford e Simon D. M. White dell'Università dell'Arizona hanno cominciato a realizzare simulazioni al calcolatore basate sul modello a materia oscura fredda, ed è parso che queste simulazioni riproducano con notevole accuratezza la distribuzione delle galassie osservata. Il modello proposto e sostenuto da questa «banda dei quattro», come è stata chiamata, è diventato la spiegazione meglio accreditata della formazione delle galassie.

Il fatto che non fosse mai stata rivelata neanche una particella stabile di materia oscura fredda era considerato una sfida per il futuro, più che un difetto esiziale: in fondo i fisici delle particelle, nel loro impegno per realizzare una teoria di grande unificazione, avevano inventato tutta una schiera di particelle ipotetiche per mediare l'interazione tra le varie forze. Si poteva quindi pensare - ipotizzavano questi cosmologi - che una di esse esistesse ancora come residuo dell'era GUT, quella della grande unificazione e in quantità sufficiente a giustificare la massa mancante.

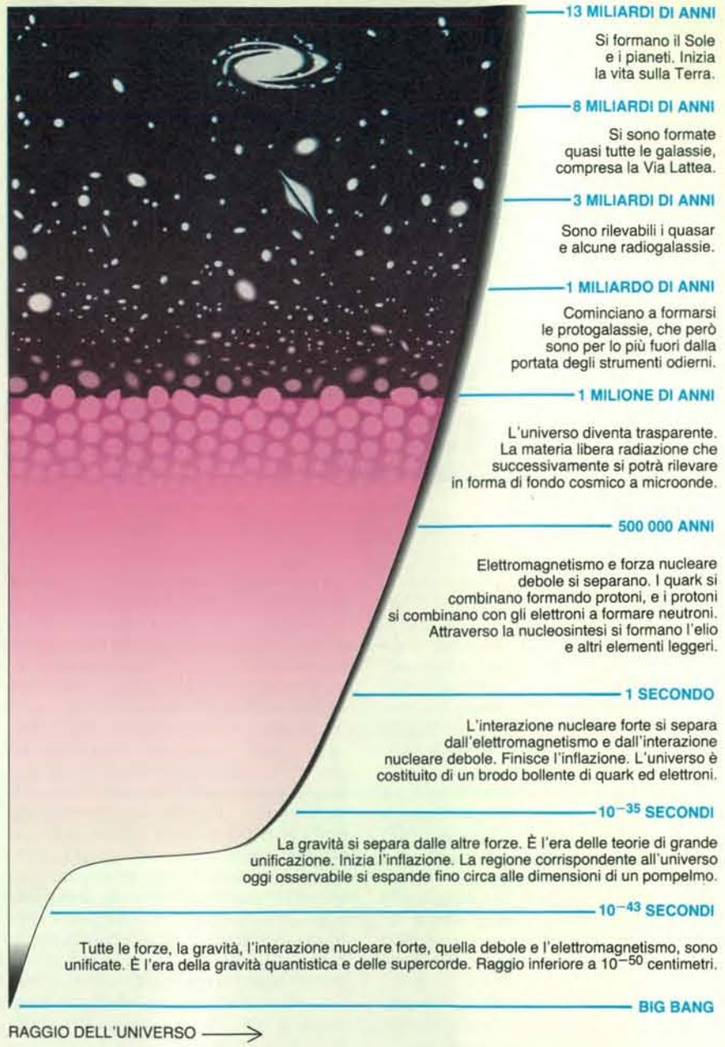
Durante un intervento al convegno, Michael S. Turner dell'Università di Chicago, noto per i suoi lucidi fantasiosi, menziona come candidati a questo ruolo i «residui di cosmologi da 100 chili»; il lucido che proietta raffigura oggetti allungati con le diciture «Dave» e «Rocky», i soprannomi con cui sono noti nell'ambiente universitario di Chicago Schramm ed Edward W. Kolb. Due candidati più plausibili sono gli assioni e le particelle supersimmetriche leggere. Gli assioni sono particelle relativamente leggere, inventate per risolvere alcuni problemi in certe teorie di grande unificazione, mentre le particelle supersimmetriche leggere sono il frutto della teoria supersimmetrica, secondo la quale tutte le particelle identificate fino a oggi posseggono un gemello che interagisce solo per via debole.

I fisici delle particelle hanno unito ben volentieri le loro forze a quelle dei cosmologi nel tentativo di identificare un candidato per la materia oscura fredda. John Ellis del CERN sostiene che trovare una particella di questo tipo rappresenterebbe un ampliamento gratificante della rivoluzione copernicana: «Non solo non ci troveremmo al centro dell'universo», dice, e sembra che la barba arruffata e gli occhi stralunati stiano a indicare un disinteresse per le preoccupazioni terrene, «ma non saremmo neanche fatti della stessa roba della maggior parte dell'universo». Ma gli esperimenti compiuti al CERN e in altri centri - afferma Krauss - hanno già escluso alcuni candidati (come un neutrino freddo di grande massa), e hanno indicato che altri aspiranti, come le particelle supersimmetriche leggere, occupano forse livelli di energia fuori dalla portata degli acceleratori attuali.

Il modello a materia oscura fredda è afflitto anche da altri problemi. Uno, per amara ironia, deriva proprio dalla radiazione di fondo a microonde che fino a oggi aveva offerto tante soddisfazioni ai cosmologi. La sua uniformità in tutte le direzioni, l'isotropia, ne costituisce da sempre una caratteristica qualificante, indicativa dell'origine primordiale. Ma il modello a materia oscura fredda prevede che le fluttuazioni a cui si deve la nascita delle galassie e delle strutture più grandi abbiano lasciato un'impronta, in forma di variazioni di temperatura locali, sul fondo di radiazione.

Questa impronta ancora non si trova.

IL BIG BANG: UNA CRONOLOGIA



QUESTA MAPPA DI DUE MILIONI di galassie, realizzata da astronomi dell'Università di Oxford, mostra addensamenti su scale molto maggiori di quelle previste dal modello a materia oscura fredda. L'immagine grande è costituita da migliaia di pixel, elementi d'immagine, ciascuno dei quali corrisponde a una lastra fotografica di una diversa zona del cielo. Ciascun pixel è bianco, grigio o nero a seconda che la lastra contenga rispettivamente più di venti galassie, da una a venti galassie o nessuna galassia.

Il *Cosmic Background Explorer* (COBE), messo in orbita l'anno scorso, ha scandagliato la radiazione fino a una risoluzione di una parte su 10 000 senza individuare alcuna fluttuazione di intensità locale. Questi valori sono già al limite della compatibilità con il modello a materia oscura fredda, e George F. Smoot dell'Università della California a Berkeley e membro del gruppo del COBE lascia intendere che nuovi dati forniti dal satellite indicheranno una uniformità della radiazione a risoluzioni ancora più fini. Neppure le misurazioni del fondo a microonde compiute da un osservatorio situato al Polo Sud, che scandaglia una regione di cielo più piccola di quella accessibile al COBE, ma a una risoluzione superiore, sono riuscite a individuare disomogeneità.

Nel frattempo un numero via via crescente di osservazioni continua a mostrare che la distribuzione delle galassie è tutto fuorché uniforme. A metà degli anni ottanta un gruppo di studio guidato da Margaret J. Geller e John P. Huchra dello Harvard-Smithsonian Center for Astrophysics ha iniziato a realizzare carte della posizione delle galassie analizzandone lo spostamento verso il rosso (che, come Hubble aveva dimostrato, è proporzionale alla distanza); si è così riscontrato che esse tendono a disporsi in filamenti e strisce enormi. Un altro gruppo di quello stesso centro ha trovato indizi dell'esistenza di una vasta regione vuota nella direzione della costellazione di Boote, primo di una lunga serie di buchi a essere identificati. Secondo il gruppo «dei sette samurai» (gli astronomi hanno una singolare predilezione per questi coloriti soprannomi) sembra che la Via Lattea e molte galassie circostanti stiano cadendo verso una grande massa non identificata, il Grande Attrattore.

Due risultati riferiti nell'ultimo anno hanno reso la situazione ancora più grave: la Geller e Huchra affermano di avere scoperto la struttura più grande mai identificata fino a oggi, la Grande Parete, una fascia di galassie dello spessore di almeno 500 milioni di anni luce, mentre Thomas J. Broadhurst dell'Università di Londra e colleghi hanno studiato due profondi e sottili «pennelli» di spazio, scoprendo che fino a una distanza di vari miliardi di anni luce si alternano densi grumi di galassie e spazi vuoti anche di 400 milioni di anni luce. La Geller, tra gli altri, sostiene che a spiegare queste osservazioni non serve né il modello a materia oscura fredda, né alcun'altra teoria: per permettere alla gravità di formare strutture come la Grande Parete a partire dalle piccole fluttuazioni di densità compatibili con la radiazione a microonde non c'è stato abbastanza tempo, punto e basta.

All'incontro in Svezia non ci sono rappresentanti di questi due gruppi, e i loro risultati ne escono con le ossa rotte. Passando in rassegna i dati sui pennelli di spazio, Schramm subisce interruzioni continue. «Sto soltanto presentando i dati - esclama - non li sto interpretando!» Per Frenk i dati sulla Grande Parete e sui pennelli sono «spazzatura»; «Non si tratta di un'analisi statistica - osserva - sono solo bei discorsi.» Persino Peebles, che non è certo un sostenitore della materia oscura fredda, ammonisce a «non fidarsi delle Grandi Pareti perché l'occhio è molto bravo a riconoscere schemi regolari».

I partecipanti alla conferenza sembrano più colpiti da un rilevamento effettuato dagli astronomi di Oxford, che non hanno elaborato un grafico della distanza delle singole galassie in base allo spostamento verso il rosso, ma in compenso hanno analizzato un numero di galassie molto elevato, prendendo in esame due milioni di oggetti su una vasta fascia di cielo. Nelle carte precedenti il rilevamento riguardava qualche migliaio di galassie in una regione piuttosto stretta.

«Abbiamo trovato indizi di un addensamento su scale almeno doppie di quelle che prevede il modello della materia oscura fredda» racconta Efstathiou, uno dei partecipanti alla ricerca, e aggiunge che il modello va perciò rivisto o scartato. L'uditorio pare attonito: Efstathiou, magro ed elegante, emana razionalità e freddezza, ed è anche uno degli autori del modello. «È come se il Papa si convertisse all'ebraismo» osserva uno degli astronomi presenti.

C'è un altro insieme di osservazioni (quelle relative a oggetti estremamente lontani e quindi estremamente vecchi) che crea dei problemi. Il modello a materia oscura fredda prevede che la grande maggioranza delle galassie impieghi almeno numerosi miliardi di anni per formarsi, e quindi a distanze superiori a 10 miliardi di anni luce se ne dovrebbero trovare poche (supponendo che l'universo abbia 13 miliardi di anni). Ma oltre questo limite sono state scoperte centinaia di quasar, punti brillantissimi che emettono luce e onde radio, ritenuti da molti astronomi galassie con un nucleo iperattivo.

Dato che i quasar sono oggetti ancora poco conosciuti, si dà maggiore importanza agli oggetti che sono manifestamente galassie. Oggi gli astronomi hanno identificato una ventina di galassie lontane più di 10 miliardi di anni luce, e cinque o sei oltre gli 11. Si tratta per lo più di radiogalassie, grandi sistemi ellittici che emettono onde radio. Sulla base di queste osservazioni il modello a materia oscura fredda non è ancora da scartare completamente, afferma Kenneth C. Chambers dell'Osservatorio di Leida, che studia le galassie lontane, «ma se se ne aggiungeranno altre, il modello passerà dei brutti guai».

Gli piacciono col botto

Già prima di quest'ultima ondata di risultati sperimentali, i fisici teorici avevano cominciato a prospettare meccanismi alternativi per la creazione delle galassie e la loro organizzazione in strutture più grandi, anche se finora il modello a materia oscura fredda rimane il più condiviso. Un'ipotesi, elaborata da Alexander Vilenkin della Tufts University e da altri, è quella delle corde cosmiche. Le corde cosmiche sono state paragonate ai difetti di struttura cristallina che si sviluppano tra le zone di un laghetto che geli in momenti diversi; analogamente, durante il raffreddamento dell'universo primordiale, si sarebbero potuti formare dei difetti nella topologia dello spazio. Questi difetti potrebbero avere avuto una massa sufficiente a formare le galassie, o perfino gli ammassi. Gli ultimi studi sulle corde inducono però a ritenere che queste sarebbero troppo piccole e troppo poco stabili per dar vita a delle galassie. «Il quadro non è promettente come qualche tempo fa», afferma Thomas W. B. Kibble dell'Università di Londra, un pioniere dello studio dei difetti topologici.

Un destino analogo è toccato alle proposte di Jeremiah P. Ostriker di Princeton e di altri su esplosioni avvenute nelle prime fasi dell'evoluzione dell'universo. Secondo questi ricercatori, a stimolare la formazione delle galassie avrebbero potuto provvedere stelle in esplosione, ossia supernove, mentre l'esplosione delle corde cosmiche avrebbe spinto le galassie in ammassi: «A Jerry - osserva Frenk - piacciono con il botto.» Ma molti cosmologi, Ostriker compreso, pensano che esplosioni abbastanza potenti, tali da determinare la formazione di grandi ammassi e vuoti, avrebbero dovuto lasciare un'impronta percepibile nella radiazione di fondo a microonde.

C'è un'altra teoria sulla formazione delle strutture che sta guadagnando in popolarità. Si tratta della teoria delle intersezioni, elaborata negli ultimi due anni da un giovane docente di Princeton già sostenitore delle corde cosmiche, Neil G. Turok, e da altri. Le intersezioni assomigliano alle corde cosmiche tranne per il fatto che sono difetti dello spazio localizzati e a forma di nodo e, per questo, più facili da simulare in modelli al calcolatore. Per Kibble si tratta di oggetti «piuttosto interessanti anche se - avverte - pure le corde cosmiche e le esplosioni, all'inizio, sembravano promettenti». In generale, ricorda, «più a fondo si guarda, più problemi si trovano».

Se il modello a materia oscura fredda se l'è cavata meglio di quasi tutti gli altri è perché i suoi sostenitori lo hanno reso più flessibile ammettendo, per esempio, che le galassie diano un'immagine di parte, imprecisa, della distribuzione vera della massa dell'universo; ciò permette loro di ridistribuire la materia come preferiscono. Il guaio di questa idea è che aumenta il numero dei parametri liberi, le variabili interne al modello che si possono regolare a volontà. I parametri liberi rendono più facile adattare il modello alle osservazioni, afferma Peebles, rendendolo però «meno capace di previsioni e meno soddisfacente».

Sono state proposte anche modifiche della teoria inflazionaria capaci di dar vita a strutture di scala maggiore. Alexei A. Starobinski dell'Istituto di fisica teorica L. D. Landau di Mosca ha proposto che l'inflazione si verifichi in diversi stadi o passi, mentre Paul J. Steinhardt dell'Università della Pennsylvania ne ha formulato una versione, l'inflazione estesa, nella quale l'intensità della gravità varia.

Il più inventivo tra gli studiosi di teoria dell'inflazione è forse Andrei D. Linde, un fisico sovietico che due anni fa è stato al CERN e l'autunno scorso alla Stanford University. Linde è una specie di fantasista capace di esordire durante un cocktail all'aperto tenuto in occasione del congresso con un salto mortale all'indietro, di spezzare in due una pietra con un colpo di karatè e di fare giochi di prestigio. Tiene un fiammifero in equilibrio sulla punta di un dito apparentemente immobile e lo fa fremere e saltare come un burattino attaccato a un filo invisibile. Ai colleghi trasecolati che gli chiedono il segreto risponde, con profondo accento russo, «sono fluttuazioni quantistiche».

Linde è famoso anche per l'abilità con cui aiuta la teoria inflazionaria a tirare avanti. Già nel 1982, è riuscito a risolvere con uno sfoggio di bravura i problemi teorici che affliggevano la formulazione originaria di Guth. Da allora Linde ne ha proposto una serie innumerevole di versioni, con assioni, corde cosmiche, neutrini, intersezioni e altri concetti ipotetici. Come se non bastasse, a differenza della maggior parte dei cosmologi, non si fa scrupoli neppure di invocare il principio antropico: «Le possibilità strane sono molte» sostiene.

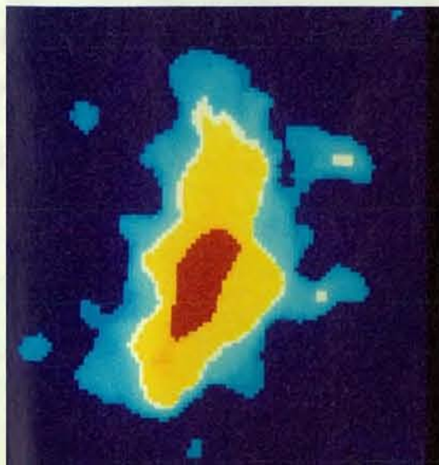
Eppure alcuni cosmologi si sentono sempre più insoddisfatti per la variabilità della teoria inflazionaria. Durante una discussione sull'inflazione estesa, Peebles appare sempre più a disagio finché sbotta: «Ma così intorno a quella membrana ti ritroverai un anello di buchi neri!». E più tardi, sviluppando ulteriormente la sua critica: «Siamo sicuri che la "solidità" della teoria inflazionaria sia un pregio? Qui sembra voglia dire capacità di adattarsi a qualsiasi dato.» E Frank Wilczek, un fisico delle particelle dell'Institute for Advanced Study di Princeton, gli fa eco: «Ci servono modelli inflazionari più specifici, verificabili. Non possiamo permettere ad Andrei di andare avanti così.»

La più grande sciocchezza di Einstein

Chi voglia scommettere sul risultato di questa giostra di teorie farà bene a tenere presente quella specie di jolly che è la costante cosmologica. Il concetto fu ideato e battezzato da Einstein agli inizi del secolo per rappresentare l'energia posseduta dal vuoto perfetto. Einstein aveva ipotizzato che occorresse attribuire al vuoto una certa energia, capace di generare una specie di pressione «centrifuga» nello spazio, per controbilanciare l'attrazione «centripeta» della gravità in un universo che si credeva statico, cioè non in espansione. Dopo la scoperta dell'espansione dell'universo, Einstein abbandonò l'idea (dicendo, pare, che era stata la sua «sciocchezza più grande») ma questa di tanto in tanto riprende a fare capolino in cosmologia.

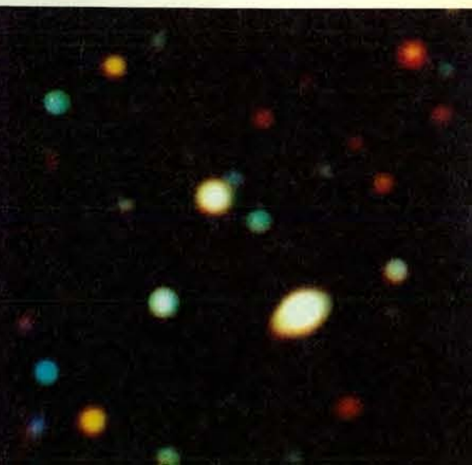
Cowie, per esempio, sostiene che una costante cosmologica potrebbe contribuire a spiegare le immagini di galassie nello spazio profondo da lui recentemente ottenute e da cui sembra che cinque miliardi di anni fa le galassie fossero varie volte più numerose di oggi. Che cosa ne è successo?

Lo scozzese fa osservare che molte potrebbero essersi fuse o essersi affievolite fino a diventare invisibili, ma preferisce pensare a un'altra spiegazione: proprio la costante cosmologica. Una piccola energia del vuoto distorcerebbe la nostra prospettiva dello spazio facendocene sottovalutare il volume su scale molto grandi. L'effetto potrebbe spiegare perché, nelle immagini dello spazio profon-



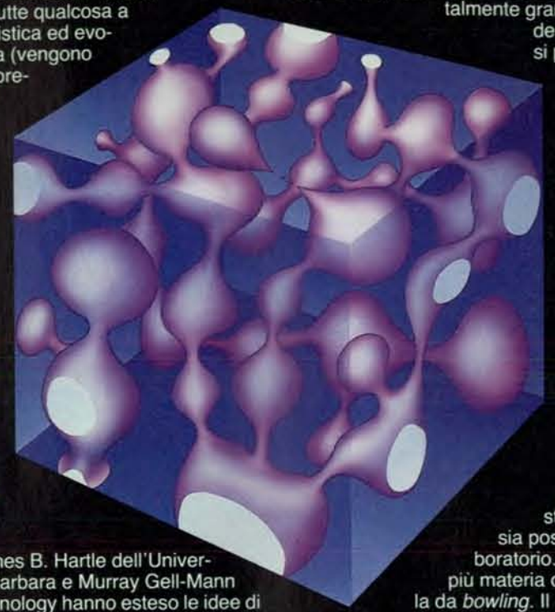
LE IMMAGINI DELLE GALASSIE LONTANE SFIDANO I COSMOLOGI

La galassia più lontana fino a oggi identificata (*a sinistra*) è stata scoperta da Kenneth C. Chambers e George K. Miley dell'Osservatorio di Leida. È lontana quasi 12 miliardi di anni luce. I cosmologi incontrano notevoli difficoltà a spiegare come abbiano fatto a formarsi galassie così poco tempo dopo il big bang. L'immagine di un campo di galassie lontano circa cinque miliardi di anni luce (*a destra*), realizzata da Lennox L. Cowie dell'Università di Hawaii e altri, fa nascere un ulteriore problema: sembra che le galassie siano molto più numerose che nelle regioni vicine; come si spiega?



UN UNIVERSO QUI, UNO LÀ...

I cosmologi continuano a rimettere tutte le cose al posto che loro compete. Non contenti di averci detto che il Sole non è che una tra i miliardi di stelle della Galassia, che a sua volta è solo una dei miliardi di galassie dell'universo, alcuni cosmologi vorrebbero farci credere anche che il nostro universo non sia che uno in una moltitudine infinita di mondi. Di teorie sui «multiversi» ne esistono parecchie; pur avendo tutte qualcosa a che vedere con la fisica quantistica ed evocando immagini simili a questa (vengono alla mente dei grumi di lava), presentano, però, differenze di grandissima rilevanza.



La teoria multi-mondi. È stata proposta da Hugh Everett III della Princeton University più di 30 anni fa per risolvere alcuni problemi relativi al ruolo dell'osservatore in fisica quantistica. In base alla fisica quantistica una particella, per esempio un elettrone, sembra seguire più traiettorie contemporaneamente, ma se un fisico la osserva, la trova lungo una sola di esse. Preoccupato dell'arbitrarietà con cui le altre vengono eliminate, Everett avanzò l'ipotesi che l'elettrone, in realtà, segua contemporaneamente tutte le traiettorie, ma ciascuna in un universo differente.

La teoria multi-storie. James B. Hartle dell'Università della California a Santa Barbara e Murray Gell-Mann del California Institute of Technology hanno esteso le idee di Everett a tutto l'universo. Immediatamente dopo il big bang, chiarisce Hartle, l'universo era talmente piccolo da potersi considerare alla stregua di una particella subatomica in moto su diverse traiettorie. Hartle preferisce il termine «storie» a quello di «mondi» e, contrariamente a Everett, considera le traiettorie alternative «potenzialità», piuttosto che realtà.

L'inflazione caotica. Il fisico sovietico Andrei D. Linde ipotizza che l'universo, quando aveva circa 10^{-35} secondi, fosse simile a una schiuma caotica. Le diverse regioni, che avevano proprietà fisiche differenti, subirono l'inflazione, un'ondata di crescita breve ma intensissima, in momenti distinti. Dopo l'inflazione, queste regioni sono rimaste separate da distanze talmente grandi da trovarsi molto al di fuori della sfera d'influenza reciproca; si potrebbe addirittura parlare, in pratica, di una pluralità di cosmi separati.

I cunicoli. Stephen W. Hawking dell'Università di Cambridge, Sidney R. Coleman della Harvard University e altri propongono che anche lo spazio-tempo possa passare per effetto tunnel da un punto all'altro proprio come fanno gli elettroni. L'effetto tunnel nel caso dello spazio-tempo crea una sorta di «cunicoli», i quali possono condurre sia ad altri punti dello stesso universo, sia a vicoli ciechi, gli universi *baby*, sia ad altri universi grandi come il nostro.

L'universo fai da te. Alan H. Guth del Massachusetts Institute of Technology pensa che sia possibile creare un universo in laboratorio. L'impresa richiederebbe poca più materia di quanta ne contenga una palla da *bowling*. Il difficile sarebbe comprimere la materia fino a densità analoghe a quelle di un buco nero e poi riuscire in qualche modo a farla espandere, proprio come è successo all'universo quando si è verificato il big bang. «Mi piace considerarlo un problema tecnico - osserva Guth - risolvibile forse da qualche civiltà del futuro.»

do, le galassie sembrino così ravvicinate.

Analogamente, la costante potrebbe far sparire una discrepanza imbarazzante fra l'età degli ammassi globulari, fitti assembramenti di stelle che ruotano intorno al centro della Via Lattea, e quella dell'universo. L'analisi spettrale delle stelle degli ammassi globulari fa attribuire loro un'età superiore ai 15 miliardi di anni, mentre le stime della distanza delle galassie e della loro velocità di recessione fanno pensare che l'universo abbia solo 13 miliardi di anni. L'esistenza di una piccola energia del vuoto porterebbe a sottovalutare il volume totale dell'universo e quindi la sua età.

Infine, l'esistenza di una costante cosmologica potrebbe spiegare perché l'universo sembra così piatto anche se la densità di massa apparente resta molto al di sotto del valore critico: dato che massa ed energia sono in un certo senso equivalenti, l'energia del vuoto potrebbe fungere da massa mancante. «Se l'intervento di Len Cowie vi ha sconcertato - sottolinea Peebles ai colleghi riuniti in Svezia - vuol dire che siete stati proprio attenti.»

La costante cosmologica, però, non gode di un grandissimo successo tra i cosmologi. L'obiezione più diffusa è che sia «brutta»: introduce un ulteriore parametro libero nei modelli, togliendo loro in capacità previsionali e in eleganza. Per di più, nessuno ha ancora saputo spiegare neppure perché la natura dovrebbe preferire proprio una costante cosmologica positiva invece che negativa; quanto a indicarne un valore preciso, non se ne parla neanche.

In effetti i calcoli di Sidney R. Coleman, un fisico della Harvard University, inducono a ritenere che la costante cosmologica dovrebbe valere esattamente zero. La spiegazione è un po' complicata: occorre tener presente che il fenomeno quantistico detto effetto tunnel, quello che permette alle particelle subatomiche di passare istantaneamente da un punto all'altro, varrebbe anche per lo spazio e il tempo. L'effetto tunnel dello spazio-tempo, secondo Coleman, genera una sorta di cunicoli che collegano il nostro universo con altri in modo da rendere nulla l'energia del vuoto.

La cosmologia quantistica

L'affermazione di Coleman sulla costante cosmologica è uno dei risultati più concreti della cosmologia quantistica, una branca della fisica quanto mai inconsueta che studia i primi 10^{-43} secondi di vita dell'universo quando questo era tanto piccolo, denso e caldo che gli effetti quantistici e quelli della gravità, appartenenti di norma a regni separati, dovevano essere strettamente legati. Anche le teorie di grande unificazione si fermano molto prima di quest'epoca, ma alcuni fisici sperano di illuminarla elaborando una teoria quantistica della gravità, altrimenti detta «teoria del tutto», che unisca tutte le forze della natura.

Le discussioni di cosmologia quantistica sanno un po' di fantascienza, e gli scienziati di indole più empirista le considerano con un misto di scetticismo e di perplessità. «Gli astronomi si addormentino pure tutti, adesso - avverte Coleman all'inizio del suo intervento sui cunicoli - ma Alan deve stare sveglio.» Eppure i cultori di questa disciplina, come John P. Preskill del California Institute of Technology, pensano che i loro sforzi potrebbero condurre a quel grande balzo in avanti teorico del quale la cosmologia sembra avere sempre più bisogno. «Alla fine - dice Preskill - queste idee potrebbero aiutarci a capire la struttura di grande scala.»

Uno degli esploratori più arditi della cosmologia quantistica è Stephen W. Hawking dell'Università di Cambridge. Colpito da una malattia degenerativa del sistema nervoso più di 20 anni fa, Hawking deve ricorrere a una cannula tracheale per respirare e riesce a muovere solo due dita della mano sinistra, ma tanto basta. Giace nella sedia a rotelle e premendo un pulsante quasi incastrato nel palmo della mano costruisce le frasi sul video di un calcolatore che gli sta di fronte. Un sintetizzatore vocale poi pronuncia le frasi per lui con una voce profonda e autorevole stranamente incongrua con il suo aspetto.

Hawking, il cui lavoro sui cunicoli nello spazio-tempo ha fatto da trampolino di lancio per l'attacco alla costante cosmologica da parte di Coleman, è convinto che questi oggetti possano condurre prima o poi a una teoria della gravità quantistica. Altri, come Gerard 't Hooft dell'Università di Utrecht, secondo il quale i cunicoli violano il fondamentale principio di causalità, ne dubitano: «Preferirei di gran lunga vedere una teoria dalla quale fossero banditi» afferma.

I giorni della cartaccia sono finiti

Un altro modo di accostarsi allo studio della gravità quantistica è quello che coinvolge le supercorde. Questi oggetti, da non confondere con le corde cosmiche, sono particelle ipotetiche dotate di lunghezza, ma non di larghezza, attorcigliate in ben 26 dimensioni. Chi ne propone lo studio sostiene che questi «riccioli» riescono a spiegare il comportamento di tutte le particelle elementari. Altri fisici notano però che la teoria matematica di questi oggetti, così astrusa, deve riuscire a dar prova di sé a energie inferiori, nel dominio accessibile agli acceleratori di particelle, prima di poter essere applicata alla creazione dell'universo. «La teoria delle corde è ancora promettente - osserva con sarcasmo Wilczek dell'Institute for Advanced Study - molto promettente, sempre promettente.» E questo vale secondo lui per tutta la cosmologia quantistica: «Sono tentativi validi, ma potrebbero essere sbagliati dal primo all'ultimo.»

Turner loda questi tentativi teorici, ma prevede che il futuro della cosmologia sarà dominato dalle osservazioni. Sono finiti, afferma, «i giorni della cartaccia a ruota libera», delle elucubrazioni teoriche senza la verifica dei dati. Eppure esprime la speranza che le teorie saranno utili come guida degli esperimenti: «Se ci sono solo osservazioni è botanica - dice - mentre se c'è solo teoria è filosofia.»

Turner sottolinea che, a onta della grande impressione destata da tutte le relazioni sulla struttura di grande scala presentate negli ultimi tempi, manca ancora una carta della maggior parte dell'universo osservabile. La situazione potrebbe migliorare, almeno in parte, grazie a un progetto comune delle Università di Princeton e di Chicago per la realizzazione di una carta della disposizione di un milione di galassie e di varie centinaia di migliaia di quasar entro la fine del secolo. Tra qualche anno lo *Hubble Space Telescope*, se verrà riparato, potrebbe rivelare la struttura delle galassie e dei quasar lontani, mentre un radiotelescopio attualmente in costruzione in India potrebbe spingersi ancora oltre individuando gli embrioni delle galassie.

Altri esperimenti potrebbero rivelare la materia oscura. Un gruppo di studio diretto da J. Anthony Tyson degli AT&T Bell Laboratories sta sondando lo spazio profondo alla ricerca di quasar e galassie lontane la cui immagine sia stata distorta, cioè piegata ad arco, ad anello o scissa in più immagini, da lenti gravitazionali costituite da materia oscura. Pierre Sikivie dell'Università della Florida e altri vogliono trasformare un grande magnete del Lawrence Livermore National Laboratory, originariamente costruito per esperimenti di fusione, in un rivelatore di assioni, e vari laboratori statunitensi ed europei stanno progettando strumenti capaci di rivelare particelle supersimmetriche leggere.

A breve termine, afferma Turner, la prova del fuoco per tutti i modelli verrà però dalla radiazione di fondo a microonde: se il COBE o l'osservatorio al Polo Sud non riuscissero abbastanza presto a trovare delle fluttuazioni «tremerebbero le fondamenta stesse della cosmologia».

E se così fosse? Potrebbe cadere perfino il modello del big bang? Se al convegno svedese nessuno ha voluto giocare al «tiro al big bang», altrove gli scienziati mostrano meno scrupoli. Lo svedese Hannes Alfven, premio Nobel per la fisica nel 1970, ha richiamato l'attenzione del mondo scientifico su un modello che fa a meno del big bang e postula che sia l'elettromagnetismo, non la gravità, il fattore dominante che dà forma alla materia nell'universo. Halton C. Arp, del Max Planck Institut für Astrophysik, sostiene invece che lo spostamento verso il rosso delle galassie potrebbe non dipendere dall'espansione dell'universo.

Schramm, il più acceso difensore del big bang, continua a sostenere che la teoria sopravviverà a questi attacchi, anche se non si troverà nessuna fluttuazione nella radiazione di fondo a microonde. Questa evenienza condannerebbe quasi tutte le spiegazioni principali della formazione delle galassie, compresa la materia oscura fredda ma, dice, aiuterebbe «modelli più inconsueti e curiosi», e ne ricorda uno secondo il quale l'universo avrebbe sviluppato i semi delle strutture di grande scala in una transizione di fase verificatasi molto tardi, dopo la liberazione del fondo a microonde.

Questi discorsi preoccupano Turok, il giovane fautore delle intersezioni, il quale teme che via via che le osservazioni rivelano un'immagine sempre più particolareggiata dell'universo, i cosmologi si trovino costretti a ricorrere a modelli sempre più complicati e abborracciati, come quello che Tolomeo aveva realizzato per descrivere la rotazione del Sole e dei pianeti intorno alla Terra. In quel caso Turok prospetta l'idea di abbandonare la cosmologia e passare a un altro ramo, magari la fisica della materia condensata. «Forse i problemi che la cosmologia si è assegnata si riveleranno troppo difficili da risolvere scientificamente; - osserva - in fondo ci vuole un bel coraggio per immaginare che l'universo si possa descrivere con una teoria semplice.»

A sentire le preoccupazioni di Turok, Peebles sorride: ricorda che quando aveva dato il primo sguardo alla cosmologia, verso la fine degli anni cinquanta, la mancanza di sostanza lo «disgustava». «Da allora la disciplina si è fatta sempre più ricca - ricorda - e non vedo motivo perché non continui così.» Per lui la marea montante di nuove osservazioni è motivo di soddisfazione, non di dispiacere. Queste osservazioni elimineranno magari molte teorie, ma alla fine, data l'abbondanza in cosmologia di talenti e creatività, daranno vita a teorie nuove con capacità esplicative molto maggiori. Nel frattempo può darsi che i cosmologi, e tutti gli altri, si trovino costretti ad abbandonare i tentativi di capire l'universo per meravigliarsi soltanto di fronte alla sua infinita complessità e stranezza.

BIBLIOGRAFIA

GUTH ALAN H. e STEINHARDT PAUL J., *L'universo inflazionario* in «Le Scienze» n. 191, luglio 1984.

SCHRAMM DAVID N. e STEIGMAN GARY, *Gli acceleratori di particelle verificano la teoria cosmologica* in «Le Scienze» n. 240, agosto 1988.

GULKIS SAMUEL, LUBIN PHILIP M., MEYER STEPHAN S. e SILVERBERG ROBERT F., *Il satellite* Cosmic Background Explorer in «Le Scienze» n. 259, marzo 1990.

KRAUSS LAWRENCE M., *The Fifth Essence: The Search for Dark Matter in the Universe*, Basic Books, 1990.

LIGHTMAN ALAN e BRAWER ROBERTA, *Origins: The Lives and Worlds of Modern Cosmologists*, Harvard University Press, 1990.

L'ANGOLO MATEMATICO

di Ian Stewart

Il teorema di Natale di Fermat

Era la vigilia di Matnatale. Mentre le due lancette della pendola si spostavano inesorabilmente sulla verticale, un numero perfetto di rintocchi si diffuse nella gelida stanzetta. Bob Scratchit asciugò il registro con un foglio di carta Möbius (una carta con un lato solo in modo da non far passare l'inchiostro), lo chiuse e lo rimise sullo scaffale. Iniziava il suo giorno di ferie annuale! Quattro fattoriale ore senza lavoro! Indossò il soprabito, si avvolse al collo la sciarpa, così logora da avere dimensione frattale inferiore a 2, e chiuse la porta dietro di sé. Uscendo, passò a salutare il principale. «Buon Matnatale, signor Stooge», disse in tono cordiale.

«Hmm... Matnatale!» grugnì il vecchio. «Il negozio rimane chiuso a Matnatale, Scratchit. E lo sa che cosa vuol dire questo?»

«Vuol dire che abbiamo un intero giorno libero, signor Stooge!»

«Vuol dire, Scratchit, che domani non ci saranno clienti! Un intero giorno senza introiti! Un giorno in cui non entrerà nulla nella cassa della *Vecchia Bottega di Curiosità Matematiche*!»

Non era certo il momento adatto, ma Scratchit aveva promesso a sua moglie che l'avrebbe fatto. «Scusi, signore» iniziò con aria speranzosa.

«Che cosa c'è, ora?»

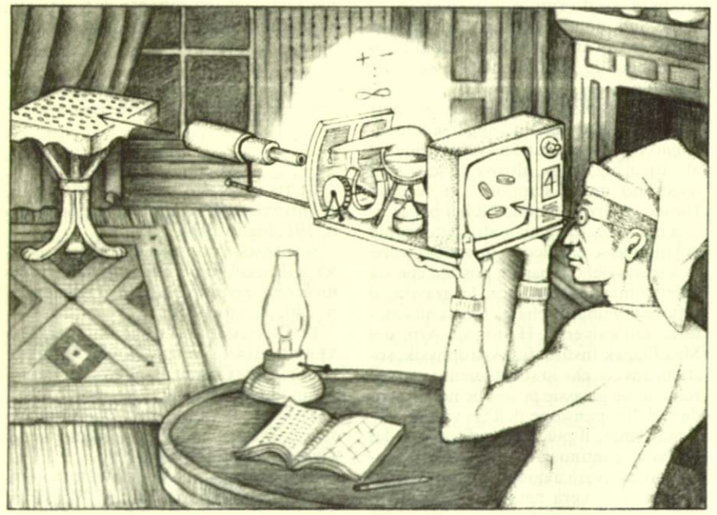
«Mi aveva promesso una gratifica natalizia, signore. Sa, è per Whiny Jim, il più piccolo, quello che continua a frignare. Appena una piccola...»

«Gratifica? Un'altra parola, Scratchit, ed è licenziato!»

Scratchit si avviò tristemente verso casa. Ma era pur sempre la vigilia di Matnatale, e presto tornò a galla la sua naturale allegria...

«Niente regali, allora?» strillò prepotentemente Whiny Jim.

«Dovremo accontentarci, figliolo» rispose Scratchit, mentre la sua allegria diminuiva esponenzialmente.



«Stooge guardò attraverso il moduloscopio le 147 monete disposte sul tavolino...»

«Io voglio un regalo! Voglio un teorema nuovo! O almeno un lemma di seconda mano! Charlie Pickens ne ha uno eccezionale! Accidenti, papà, anche una congettura sarebbe meglio di niente!»

«Mi spiace, Whiny Jim, ma il signor Stooge si tiene tutte le congetture per sé. Temo proprio di non poter tirar fuori neppure una fallacia semplice semplice. Sono completamente al verde.»

«È colpa tua, papà, se non hai alcuna ambizione. Dovresti trovarti un altro lavoro. Dall'*Onesto Pitagora* si libera un posto!»

«Whiny Jim, io non sono una persona orgogliosa, ma non scenderò mai tanto in basso da vendere triangoli rettangoli sgangherati con l'ipotenusa storta!» Scratchit si sforzò di calmarsi. «Avremo il solito pudding stantio avanzato a Pasqua, e te lo farai piacere come tutti gli altri Matnatali! Se proprio sarai fortunato, forse riuscirò a scovare un vecchio paradosso che mi ha dato tua madre. Una mano fresca di logica e sarà come nuovo!»

«Logica intuizionista?» chiese speranzoso Jim. «Non la solita roba a due valori?»

«Idea brillante, ragazzo mio!» convenne Scratchit. Whiny Jim se ne andò, temporaneamente placato, lasciando Scratchit a chiedersi freneticamente da dove diavolo poteva tirar fuori un'affermazione la cui veridicità o falsità non potesse essere determinata.

Nel suo polveroso appartamento all'altro capo della città Ebeneezer Stooge si sistemò per benino tra le coperte. «Matnatale non sarà tempo sprecato», pensò. «Darò un'occhiata ai registi per trovare qualche altro modo di fregare il fisco...»

Lo svegliò un vento gelido che scompigliava le tende e faceva sbattere i vetri. Saltò giù dal letto per chiudere la finestra... ma era già chiusa. Allora da dove veniva il vento?

«Ebeneezerr» sussurrò una voce cupa. Stooge schizzò di nuovo a letto e si rintanò tra le coperte.

«Chi... chi sei?»

«Io» tuonò la voce «sono il Fantasma dei teoremi del passato. Sono venuto per portarti via con me, Stooge!» Il fantasma tese la mano eterea e Stooge la afferrò con riluttanza.

Si trovò all'improvviso in una stanza rivestita di legno. Un uomo in vestaglia nera stava scrivendo con una penna d'oca.

«Dove ci troviamo?»

«In Francia. È il Natale di esattamente 350 anni fa.»

«Chi è quello strano tipo?»

«Il gentiluomo con la parrucca, Stooge, è il grande matematico Pierre de Fermat, che si occupò di teoria dei numeri ed è ricordato soprattutto per il suo "ultimo teorema", non ancora dimostrato. Sta scrivendo al suo amico Marin Mersenne una lettera in cui parla di una scoperta straordinaria. Lo sappiamo per certo perché quella lettera esiste ancora, ed è datata 25 dicembre 1640.»

«E di che scoperta si tratta?»

«È il "teorema di Natale di Fermat". Alcuni numeri primi possono essere espressi come somma di due quadrati perfetti: per esempio, $5 = 1 + 4 = 1^2 + 2^2$ oppure $13 = 4 + 9 = 2^2 + 3^2$. Altri numeri primi, come 3 o 11, non possono essere espressi nello stesso modo. Fermat scoprì come distinguere gli uni dagli altri.»

Stooge tirò fuori un taccuino tutto gualcito e iniziò a calcolare; gli ci volle poco a risolvere il problema per tutti i numeri primi fino a 100 (*si veda la tabella in questa pagina*).

«Riesci a vedere lo schema?» chiese il Fantasma dei teoremi del passato. Stooge scosse il capo. «Visto che è Matnatale, ti darò due suggerimenti. Il primo è di ignorare il numero due, che è anomalo.» (Lo è sempre, in quanto unico numero primo pari!) «Il secondo è di osservare il resto della divisione per 4. Tutti i numeri primi sono uguali a un multiplo di 4 più 1 o a un multiplo di 4 più 3; ossia, sono della forma $4k + 1$ o $4k + 3$. Per esempio $5 = 4 \times 1 + 1$ è della forma $4k + 1$.»

Stooge aggiunse una nuova colonna alla sua tabella, per indicare se un numero primo era della forma $4k + 1$ o $4k + 3$, e subito comprese quale fosse lo schema. «Sembra che i numeri primi che possono essere espressi come somma di due quadrati siano quelli della forma $4k + 1$» disse con aria sorpresa. «A parte 2 che è anomalo, naturalmente.»

«Ottimo. Ora, Fermat non si è limitato a intuirlo: lo ha dimostrato. O almeno, ha delineato un metodo per dimostrarlo.» Mentre il Fantasma dei teoremi del passato svaniva, Stooge lo sentì ancora: «Leonhard Euler lo dimostrò effettivamente, nel 1754.»

Ebeneezer Stooge si ritrovò nella sua gelida camera da letto. Cercò di dormire, ma lo strano teorema di Fermat continuava a ronzargli nella testa. Numeri primi, somme di due quadrati, resti della divisione per quattro... maledizione! Si girò, si rigirò, provò a fare uno spuntino, ma il sonno non voleva venire.

In una zona molto più povera della città, anche Bob Scratchit si rivoltava nel suo letto, chiedendosi come poteva trovare in tempo il regalo per Whiny Jim.

Infine Stooge riuscì a sprofondare in un sonno agitato, ma dopo non più di $29 = 4 + 25$ secondi, un terribile ululato e un tonfo fragoroso lo ridestarono. Parson Snows aveva forse lanciato di nuovo il vaso da notte contro il gatto della vedova Kleene? No, il rumore era venuto dall'interno ed Ebeneezer iniziò a tremare di paura: una forma luminosa si stava materializzando davanti a lui.

«Io - tuonò la forma - sono il Fantasma delle intuizioni future».

«Al diavolo! Fai pure tutto quello che ti pare, io sono troppo stanco per poter protestare.»

Il Fantasma delle intuizioni future mise una scatola sul tavolo e ordinò: «Aprila!»

All'interno Stooge trovò un oggetto che sembrava una lente da gioielliere con un quadrante. «Che cos'è?» chiese rigirandolo fra le mani.

«È un moduloscopio. Ti fa ignorare le cose che non vuoi vedere.»

«Vuoi dire, per esempio, i poveri? Ma io ci riesco benissimo lo stesso.»

«Ah, ma con un moduloscopio puoi evitare di vedere le cose in modo perfettamente costante! In particolare, se regoli il quadrante su un canale (ossia un numero) e guardi nello strumento, non vedi più tutti gli elementi che possono essere raggruppati in multipli di quel numero. Mettilo su "4" e guarda le mie mani: quante dita vedi?»

«Due».

«Esatto! Vedi le mie dita modulo 4, cioè ignorando tutti i multipli di 4; dato che 10 modulo 4 è uguale a 2, vedi solo le dita che rimangono.»

Stooge guardò attraverso il moduloscopio le 147 monete che aveva posate sul tavolino. Rabbrividì nel vederne solo 3. Allontanò dagli occhi il moduloscopio e, con sollievo, vide che erano tornate 147.

«Non distrarti! Lascia il quadrante su "4" e guarda la tua tabella dei numeri primi. Che cosa vedi?»

«Solo degli 1 e dei 3 - a parte 2, che è anomalo. Ogni numero primo che è somma di due quadrati appare come un 1 e ogni numero primo che non è somma di due quadrati appare come un 3! Ma è naturale: non è che la condizione $4k + 1$ o $4k + 3$; 1 o 3 modulo 4.» Fece una pausa. «Però continuo a non capire perché sia importante il valore del numero primo modulo 4.»

«Invece di guardare i numeri primi, guarda i quadrati!»

Stooge si riportò il moduloscopio davanti agli occhi. Ci fu un lungo silenzio. «Quello che vedo» disse finalmente «è solo l'equazione $1 = 0 + 1$ continuamente ripetuta.»

«Sì. E capisci perché?»

«Perché gli unici quadrati modulo 4 sono 0 e 1?»

«Esattamente. Se elevi al quadrato un numero pari, ottieni un multiplo di 4, che attraverso il moduloscopio ti appare come uno 0. Se elevi al quadrato un numero dispari, ottieni 1, 9, 25, 49... tutti numeri uguali a un multiplo di 4 più 1. Quindi le somme di quadrati modulo 4 sono $0 + 0 = 0$, $0 + 1 = 1$ o $1 + 1 = 2$. Che cosa manca?»

«Tre», disse Stooge.

«Proprio così. Una somma di due quadrati può essere 0, 1 o 2 modulo 4, ma mai 3. Allora un numero primo - anzi qualsiasi numero - della forma $4k + 3$ non può essere la somma di due quadrati. Ora capisci l'importanza del modulo 4, vero?» Il fantasma iniziò a svanire.

«Aspetta!» gridò Stooge disperato. «Tutto questo è giusto, ma non dimostra che tutti i numeri primi della forma $4k + 1$ sono somme di due quadrati. Dimostra solo che gli altri sicuramente non lo sono!»

La risposta giunse ormai fioca: «Hai ragione; ma quasi ci sei. Tieni vicino il moduloscopio e aspeeettaaaa...»

«Oddio, deve essere un altro fantasma» pensò Stooge. «I guai vengono sempre tre alla volta.» Si rivolse verso il soffitto e urlò: «Forza, materializzati, non possiamo perderci tutta la notte!»

«Io sono il Fant... etciù! il Fantasma delle dimostrazioni presenti. Si gela qui dentro! Ma non accendi mai il fuoco?» Il fantasma si soffiò rumorosamente il naso in un lembo di ectoplasma.

«Sei venuto qui per farmi vedere come si dimostra che ogni numero primo della forma $4k + 1$ è la somma di due quadrati?»

«Proprio così, ragazzo mio! Ci toccano sempre dei lavoracci, a noi fantasmi! Ma non perdiamo tempo, Ebeneezy Stooge. Metti il tuo quadrante su canale 17 e tutto sarà chiaro!» Con ampio gesto, il Fantasma delle dimostrazioni presenti tirò fuori un foglio di plastica suddiviso in quadrati e lo mise sul tavolo. «Ti farò vedere la dimostrazione per il numero primo 17, ma lo stesso metodo funziona per tutti. L'idea, amico mio, è smettere di pensare ai numeri primi e iniziare a pensare alle somme di quadrati. Su questo foglio speciale ci sono tutte le possibili somme di due quadrati, con $x^2 + y^2$ scritto nella colonna x e nella riga y.

NUMERO PRIMO	SOMMA DI 2 QUADRATI?	$4k + 1$ O $4k + 3$?
2	$1^2 + 1^2$	eccezione
3	no	(4 x 0) + 3
5	$1^2 + 2^2$	(4 x 1) + 1
7	no	(4 x 1) + 3
11	no	(4 x 2) + 3
13	$2^2 + 3^2$	(4 x 3) + 1
17	$1^2 + 4^2$	(4 x 4) + 1
19	no	(4 x 4) + 3
23	no	(4 x 5) + 3
29	$2^2 + 5^2$	(4 x 7) + 1
31	no	(4 x 7) + 3
37	$1^2 + 6^2$	(4 x 9) + 1
41	$4^2 + 5^2$	(4 x 10) + 1
43	no	(4 x 10) + 3
47	no	(4 x 11) + 3
53	$2^2 + 7^2$	(4 x 13) + 1
59	no	(4 x 14) + 3
61	$5^2 + 6^2$	(4 x 15) + 1
67	no	(4 x 16) + 3
71	no	(4 x 17) + 3
73	$3^2 + 8^2$	(4 x 18) + 1
79	no	(4 x 19) + 3
83	no	(4 x 20) + 3
89	$5^2 + 8^2$	(4 x 22) + 1
97	$4^2 + 9^2$	(4 x 24) + 1

Quali numeri primi sono somme di quadrati?

Guardalo attraverso il moduloscopio. Che cosa vedi?»

«Un sacco di numeri compresi tra 0 e 16, dappertutto.»

«Hmm, già, che stupido sono. Prendi questo pennarello e segna dei puntini ovunque vedi uno zero.» Stooge disegnò una strana configurazione periodica di punti (*si veda l'illustrazione in questa pagina*), stette un po' a osservarla, poi scosse la testa con aria perplessa.

«C'è uno schema nascosto» disse il fantasma. «Fammi colorare in rosso alcuni dei punti e gli altri in azzurro... Vedi qualcosa?»

«Caspita! Sono due griglie regolari, una sovrapposta all'altra!»

«Esatto! Il nome giusto per una griglia di questo genere è reticolo. I punti che hai disegnato sono tutti i punti (x, y) della colonna x e della riga y tali per cui $x^2 + y^2$ è un multiplo di 17. Ora, guarda il reticolo rosso e dimmi qual è il punto più vicino all'origine (colonna 0, riga 0).»

«Facile. È nella colonna 1, riga 4.»

«E quale multiplo di 17 è la somma di quadrati corrispondenti?»

«$1^2 + 4^2 = 17$ stesso! Ora capisco! Mi stai dicendo che il punto del reticolo rosso più vicino all'origine risolve il problema di rappresentare 17 come somma di due quadrati!»

«Sì. Anche il reticolo azzurro funziona, ma dà la soluzione rovesciata, $4^2 + 1^2$. Fai un'altra prova: questa volta metti il tuo quadrante su canale 41. Hai lo stesso tipo di risultato?»

«Sì! Guarda, abbiamo ancora due reticoli sovrapposti! Il punto più vicino all'origine nel reticolo rosso si trova in colonna 4, riga 5; e $4^2 + 5^2$ è uguale a 41!»

«Grandioso! Puoi divertirti a provare altri canali sul moduloscopio. Scegli un tuo numero primo p e segna tutte le posizioni (x, y) con $x^2 + y^2$ multiplo di p. Otterrai sempre una configurazione formata da due reticoli, anche se probabilmente non lo avresti notato se non ti fosse stato detto. Ma io sono il Fantasma delle dimostrazioni presenti, non degli esempi! Devo spiegare perché ci sono due reticoli, e perché il punto del reticolo più vicino all'origine risolve sempre il problema! Iniziamo con l'esistenza dei due reticoli, esistenza che dipende dalla radice quadrata di -1.»

«Non sapevo che ne avesse una», disse Stooge.

«Ah. Nessun numero reale può avere radice -1, quindi è stato inventato un nuovo numero i per il quale $i^2 = -1$, ed è nato così il sistema dei numeri complessi. Ma con un moduloscopio non hai bisogno dei numeri complessi.» Il fantasma scrisse qualcosa sul foglio di plastica. «Dai un'occhiata a questo attraverso il moduloscopio, ancora sul canale 17.»

Stooge guardò il foglio e lesse:

$$x^2 + y^2 = (x + 4y)\,(x - 4y).$$

Ridicolo! Ma attraverso il moduloscopio le cose sono quasi sempre ridicole, se non le si sa interpretare correttamente! Sappiamo che, nell'algebra comune, $(x + 4y)\,(x - 4y) = x^2 - 16y^2$. Ma attraverso il moduloscopio, -16 è la stessa cosa che $17 - 16$ (perché i multipli di 17 sono invisibili), cioè esattamente 1. Quindi, attraverso il moduloscopio, $x^2 - 16y^2 = x^2 + y^2$!

«I punti che hai segnato» fece notare il fantasma «sono quelli che - visti attraverso il moduloscopio - soddisfano l'equazione $x^2 + y^2 = 0$. Ma in modulo 17 questa equazione equivale a: $(x + 4y)\,(x - 4y) = 0$; di conseguenza $x = -4y$ o $x = 4y$. A ciascuno dei fattori corrisponde uno dei reticoli: quello rosso all'equazione $x = -4y$, e quello azzurro a $x = 4y$. Sempre modulo 17, naturalmente! Guarda i reticoli e controlla. Per esempio, troverai tutti i punti (4,1), (8,2), (12,3), (16,4), (20,5) eccetera, che soddisfano l'equazione $x = 4y$.

«Questa è la prima idea importante. Visto attraverso il moduloscopio, sul canale 17, il numero -1 ha una radice quadrata, e precisamente 4. In effetti, $4^2 + 1 = 17 = 0$. E questo ci porta direttamente all'esistenza dei due reticoli. Esattamente la stessa cosa avviene per qualsiasi numero primo della forma $4k + 1$; si tratta proprio di quei numeri primi, modulo i quali, -1 ha una radice quadrata. Ora sei pronto per un nuovo importante concetto?»

«Ogni reticolo è costituito da parallelogrammi identici. (Qui i parallelogrammi sono in realtà quadrati inclinati, ma per molti reticoli questo non avviene, quindi continueremo a chiamarli "parallelogrammi".) Qual è l'area di un parallelogramma di questo genere? Proviamo con qualche esempio.»

Ebeneezer scribacchiò qualcosa sul suo taccuino. «Quando $p = 17$ il parallelogramma ha area... 17 unità quadrate. E quando $p = 41$ ha area 41 unità quadrate. Suppongo che, per un generico numero primo p il parallelogramma abbia area pari a p unità quadrate.»

17	289	290	293	289	305	314	325	338	353	370	389	410	433	458	485	514	545	578
16	256	257	260	265	272	281	292	305	320	337	356	377	400	425	452	481	512	545
15	225	226	229	234	241	250	261	274	289	306	325	346	369	394	421	450	481	514
14	196	197	200	205	212	221	232	245	260	277	296	317	340	365	392	421	452	485
13	169	170	173	178	185	194	205	218	233	250	269	290	313	338	365	394	425	458
12	144	145	148	153	160	169	180	193	208	225	244	265	288	313	340	369	400	433
11	121	122	125	130	137	146	157	170	185	202	221	242	265	290	317	346	377	410
10	100	101	104	109	116	125	136	149	164	181	200	221	244	269	296	325	356	389
9	81	82	85	90	97	106	117	130	145	162	181	202	225	250	277	306	337	370
8	64	65	68	73	80	89	100	113	128	145	164	185	208	233	260	289	320	353
7	49	50	53	58	65	74	85	98	113	130	149	170	193	218	245	274	305	338
6	36	37	40	45	52	61	72	85	100	117	136	157	180	205	232	261	292	325
5	25	26	29	34	41	50	61	74	89	106	125	146	169	194	221	250	281	314
4	16	17	20	25	32	41	52	65	80	97	116	137	160	185	212	241	272	305
3	9	10	13	16	25	34	45	58	73	90	109	130	153	178	205	234	265	298
2	4	5	8	13	20	29	40	53	68	85	104	125	148	173	200	229	260	293
1	1	2	5	10	17	26	37	50	65	82	101	122	145	170	197	226	257	290
0	0	1	4	9	16	25	36	49	64	81	100	121	144	169	196	225	256	289
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17

17	0	1	4	9	16	8	2	15	13	13	15	2	8	16	9	4	1	0
16	1	2	5	10	0	9	3	16	14	14	16	3	9	0	10	5	2	1
15	4	5	8	13	3	12	6	2	0	0	2	6	12	3	13	8	5	4
14	9	10	13	1	8	0	11	7	5	5	7	11	0	8	1	13	10	9
13	16	0	3	8	15	7	1	14	12	12	14	1	7	15	8	3	0	16
12	8	9	12	0	7	16	10	6	4	4	6	10	16	7	0	12	9	8
11	2	3	6	11	1	10	4	0	15	15	0	4	10	1	11	6	3	2
10	15	16	2	7	14	6	0	13	11	11	13	0	6	14	7	2	16	15
9	13	14	0	5	12	4	15	11	9	9	11	15	4	12	5	0	14	13
8	13	14	0	5	12	4	15	11	9	9	11	15	4	12	5	0	14	13
7	15	16	2	7	14	6	0	13	11	11	13	0	6	14	7	2	16	15
6	2	3	6	11	1	10	4	0	15	15	0	4	10	1	11	6	3	2
5	8	9	12	0	7	16	10	6	4	4	6	10	16	7	0	12	9	8
4	16	0	3	8	15	7	1	14	12	12	14	1	7	15	8	3	0	16
3	9	10	13	1	8	0	11	7	5	5	7	11	0	8	1	13	10	9
2	4	5	8	13	3	12	6	2	0	0	2	6	12	3	13	8	5	4
1	1	2	5	10	0	9	3	16	14	14	16	3	9	0	10	5	2	1
0	0	1	4	9	16	8	2	15	13	13	15	2	8	16	9	4	1	0
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17

Tavola delle somme di quadrati e, a destra, come essa è vista attraverso il canale 17 del moduloscopio.

«Proprio così, anche se non mi fermerò a dimostrarlo. Ma probabilmente ti stai chiedendo perché io sia interessato all'area del parallelogramma.»

«In effetti...»

«È a causa di un teorema dimostrato da Hermann Minkowski, un matematico lituano che, agli inizi del Novecento, insegnò in Germania; egli introdusse lo spazio di Minkowski, fornendo una interpretazione geometrica ormai classica della teoria della relatività ristretta di Einstein. Minkowski ebbe un'idea brillante, anche se apparentemente semplice, a proposito dei reticoli: se l'area del parallelogramma è abbastanza piccola, allora i punti del reticolo devono essere alquanto vicini l'uno all'altro! Quindi alcuni di essi devono essere molto vicini all'origine.

«Egli dimostrò un teorema che precisava questa idea. Supponiamo di avere un reticolo, formato da parallelogrammi; e supponiamo di avere un cerchio che abbia per centro l'origine. Il teorema di Minkowski afferma che se l'area del cerchio è almeno quattro volte l'area del parallelogramma, allora qualche punto del reticolo diverso dall'origine deve trovarsi all'interno del cerchio.

«Possiamo utilizzare il teorema di Minkowski per dimostrare che il punto del reticolo più vicino all'origine risolve il problema di rappresentare p come somma di due quadrati. Prendiamo per esempio $p = 17$. Anche l'area del parallelogramma è $\sqrt{17}$. Consideriamo un cerchio di raggio leggermente superiore a $\sqrt{17}$, diciamo 5. La sua area è pari a $5^2\pi = 25\,\pi = 78{,}54$; dato che è un valore maggiore di $4 \times 17 = 68$, si può applicare il teorema di Minkowski. Mi segui, finora?»

«Pendo dalle tue labbra.»

«Per il teorema di Minkowski deve esistere un punto del reticolo, diverso dall'origine, che si trova all'interno di questo cerchio. Chiamiamolo (x, y). Allora $x^2 + y^2$ è minore o uguale al quadrato del raggio del cerchio, cioè a 25. Ma per i punti del reticolo $x^2 + y^2$ è un multiplo di 17, e un multiplo diverso da zero in quanto il punto in questione non è l'origine. Quali multipli di 17 diversi da zero sono minori o uguali a 25?»

«17 stesso» disse Stooge. «E basta.»

«Esattamente! Quindi $x^2 + y^2$ deve essere esattamente 17, e questo risolve il problema! Eccezionale, vero?» (Lo stesso metodo funziona in generale: si veda la «finestra» in questa pagina.)

«L'ardita idea di Minkowski ha dato origine a una nuova branca della matematica, detta "geometria dei numeri" dal titolo di un suo libro pubblicato nel 1896, che utilizza la geometria per studiare la teoria dei numeri; due argomenti che non si riterrebbero davvero collegati! La matematica è piena di sorprese come questa. Un'altra applicazione della geometria dei numeri è il "teorema dei quattro quadrati": ogni numero intero positivo (primo o non primo) è la somma di quattro quadrati perfetti. Ma per averne la dimostrazione dovrai aspettare fino al prossimo Matnatale, Ebeneezy.»

Stooge, la mente ormai sgombra e tranquilla, cadde in un sonno profondo. Mentre si addormentava, si ricordò del povero Bob Scratchit e, con insolita benevolenza, si ripromise di essere più gentile in futuro con lui.

Come dite? E il povero Whiny Jim? Sentite, Bob Scratchit è un vero perdente e ha tanta ambizione quanta ne ha un cetriolo di mare. Nessun fantasma che si rispetti muoverebbe un dito per lui. Quanto poi a quel pestifero marmocchio di Whiny Jim...

Ah, è Natale, e volete un lieto fine.
D'accordo, l'avrete...

«Papà, papà! Me lo hai trovato il vecchio paradosso di mamma con una mano fresca di logica intuizionista, eh, lo hai trovato, lo hai trovato? Quello con un valore di verità indeterminato?»

«Beh, veramente...» disse Scratchit. «È una cosa difficile a dirsi, figlio mio.» Consegnò il pacchettino malconcio, cercando un giro di parole per far cessare gli strilli spaccatimpani di Whiny Jim. «*Non sono sicuro se l'ho trovato o no.*»

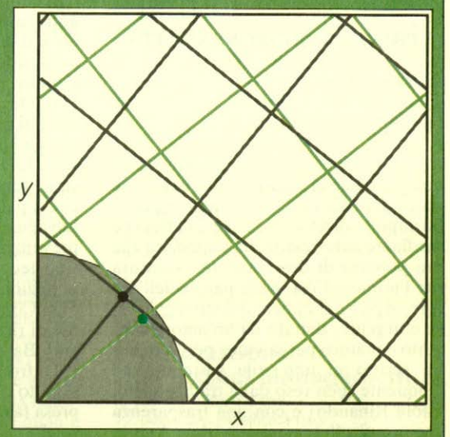
Fu un colpo di genio. O forse fu il Fantasma delle intuizioni future, appollaiato sulla sua spalla, a sussurrargli all'orecchio. Perché la faccia di Whiny Jim si illuminò come un albero di Matnatale: «Oh, papà! Lo hai trovato!»

Diciamocelo: sarebbe difficile trovare un valore di verità più indeterminato della risposta di Scratchit.

Dimostrazione di Minkowski del Teorema dei due quadrati

Supponiamo che p sia un numero primo della forma $4k + 1$, allora può essere rappresentato come somma di due quadrati.

Scegliamo uno dei due reticoli che insieme generano tutti i punti (x, y) con $x^2 + y^2$ multiplo di p. Ciascun reticolo è costituito da parallelogrammi di area pari a p unità quadrate.



Disegnamo ora un cerchio che abbia per centro l'origine e il cui raggio sia un poco maggiore di $\sqrt{p}$: diciamo $1{,}2\sqrt{p}$. L'area di questo cerchio è, pertanto, $1{,}44\,\pi\,p \approx 4{,}52p$, che è maggiore di $4p$.

Per il teorema di Minkowski, quel cerchio contiene un punto (x, y) del reticolo diverso da zero.

Per definizione di reticolo, $x^2 + y^2$ è un multiplo di p; ma dal momento che (x, y) è interno al cerchio, $x^2 + y^2 \leq 1{,}44p$.

L'unico multiplo di p diverso da zero che sia minore o uguale a $1{,}44p$ è p stesso, in quanto il più piccolo multiplo successivo è $2p$, che è troppo grande!

Quindi $x^2 + y^2 = p$ esattamente, e abbiamo dimostrato il Teorema di Fermat dei due quadrati.

BIBLIOGRAFIA

POLLARD H., *The Theory of Albebraic Numbers*, Mathematical Association of America, 1950.

KLINE MORRIS, *Mathematical Thought from Ancient to Modern Times*, Oxford University Press, 1972.

STEWART IAN e TALL DAVID, *Algebraic Number Theory*, Champman and Hall, 1987.